locus depicted in Figure 137, these strains are indicated as rlrA+. Confirming these findings, electron microscopy and negative staining detects the presence of pili extending from the surface of *S. pneumoniae*. See Figure 185. To demonstrate that the adhesin island locus was responsible for the pili, the rrgA-srtD region of TIGR 4 were deleted. Deletion of this region of the adhesin island resulted in a loss of pili expression. See Figure 186. See also Figure 235, which provides an electron micrograph of *S. pneumoniae* lacking the rrgA-srtD region immunogold stained using anti-RrgB and anti-RrgC antibodies. No pili can be seen. Similarly to that described above, a *S. pneumoniae* bacteria that lacks a transcriptional repressor, *mgrA*, of genes in the adhesin island expresses pili. See Figure 187. However, and as expected, a *S. pneumoniae* bacteria that lacks both the mgrA and adhesin island genes in the rrgA-srtD region does not express pili. See Figure 188.

These high molecular weight pili structures appear to play a role in adherence of *S. pneumoniae* to cells. *S. pneumoniae* TIGR4 that lack the pilus operon have significantly diminished ability to adhere to A549 alveolar cells in vitro. See Figure 184.

The Sp0463 (*S. pneumoniae* TIGR4 rrgB) adhesion island polypeptide is expressed in oligomeric form. Whole cell extracts were analyzed by Western blot using a Sp0463 antiserum. The antiserum cross-hybridized with high molecular weight Sp0463 polymers. See Figure 156. The antiserum did not cross-hybridize with polypeptides from D39 or R6 strains of *S. pneumoniae*, which do not contain the AI locus depicted in Figure 137. Immunogold labelling of *S. pneumoniae* TIGR 4 using RrgB antiserum confirms the presence of RrgB in pili. Figure 189 shows double-labeling of S. pneumoniae TIGR 4 bacteria with immunolabeling for RrgB (5 nm gold particles) and RrgC (10 nm gold particles) protein. The RrgB protein is detected as present at intervals along the pilus structure. The RrgC protein is detected at the tips of the pili. See Figure 234 at arrows; Figure 234 is a close up of a pilus in Figure 189 at the location indicated by *.

The RrgA protein appears to be present in and necessary for formation of high molecular weight structures on the surface of *S. pneumoniae* TIGR4. See Figure 181 which provides the results of Western blot analysis of TIGR4 *S. pneumoniae* lacking the gene encoding RrgA. No high molecular weight structures are detected in *S. pneumoniae* that do not express RrgA using antiserum raised against RrgB. See also Figure 183.

A detailed diagram of the amino acid sequence comparions of the RrgA protein in the ten *S. pneumoniae* strains is shown in Figure 148. The diagram reveals the division of the individual *S. pneumoniae* strains into the two different homology groups.

The cell surface polypeptides encoded by the *S. pneumoniae* TIGR4 AI, Sp0462 (rrgA), Sp0463 (rrgB), and Sp0464 (rrgC), have been cloned and expressed. See examples 15-17. A polyacrylamide gel showing successful recombinant expression of RrgA is provided in Figure 190A. Detection of the RrgA protein, which is expressed in pET21b with a histidine tag, is also shown by Western blot analysis in Figure 190B, using an anti-histidine tag antibody.

Antibodies that detect RrgB and RrgC antibodies have been produced in mice. See Figures 191 and 192, which show detection of RrgB and RrgC, respectively, using the raised antibodies.

In addition to the identification of these *S. pneumoniae* adhesion islands, coding sequences for SrtB type sortases have been identified in several *S. pneumoniae* clinical isolates, demonstrating conservation of a SrtB type sortase across these isolates.

Recombinantly Produced AI polypeptides

It is also an aspect of the invention to alter a non-AI polypeptide to be expressed as an AI polypeptide. The non-AI polypeptide may be genetically manipulated to additionally contain AI polypeptide sequences, *e.g.*, a sortase substrate, pilin, or E-box motif, which may cause expression of the non-AI polypeptide as an AI polypeptide. Alternatively the non-AI polypeptide may be genetically manipulated to replace an amino acid sequence within the non-AI polypeptide for AI polypeptide sequences, *e.g.*, a sortase substrate, pilin, or E-box motif, which may cause expression of the non-AI polypeptide as an AI polypeptide. Any number of amino acid residues may be added to the non-AI polypeptide or may be replaced within the non-AI polypeptide to cause its expression as an AI polypeptide. At least 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 50, 75, 100, 150, 200, or 250 amino acid residues may be replaced or added to the non-AI polypeptide amino acid sequence. GBS 322 may be one such non-AI polypeptide that may be expressed as an AI polypeptide.

GBS Adhesin Island Sequences

The GBS AI polypeptides of the invention can, of course, be prepared by various means (*e.g.* recombinant expression, purification from GBS, chemical synthesis *etc.*) and in various forms (*e.g.* native, fusions, glycosylated, non-glycosylated *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other streptococcal or host cell proteins) or substantially isolated form.

The GBS AI proteins of the invention may include polypeptide sequences having sequence identity to the identified GBS proteins. The degree of sequence identity may vary depending on the amino acid sequence (a) in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more). Polypeptides having sequence identity include homologs, orthologs, allelic variants and functional mutants of the identified GBS proteins. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affinity gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The GBS adhesin island polynucleotide sequences may include polynucleotide sequences having sequence identity to the identified GBS adhesin island polynucleotide sequences. The degree of sequence identity may vary depending on the polynucleotide sequence in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more).

The GBS adhesin island polynucleotide sequences of the invention may include polynucleotide fragments of the identified adhesin island sequences. The length of the fragment may

vary depending on the polynucleotide sequence of the specific adhesin island sequence, but the fragment is preferably at least 10 consecutive polynucleotides, (*e.g.* at least 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more).

The GBS adhesin island amino acid sequences of the invention may include polypeptide fragments of the identified GBS proteins. The length of the fragment may vary depending on the amino acid sequence of the specific GBS antigen, but the fragment is preferably at least 7 consecutive amino acids, (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). Preferably the fragment comprises one or more epitopes from the sequence. Other preferred fragments include (1) the N-terminal signal peptides of each identified GBS protein, (2) the identified GBS protein without their N-terminal signal peptides, and (3) each identified GBS protein wherein up to 10 amino acid residues (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) are deleted from the N-terminus and/or the C-terminus *e.g.* the N-terminal amino acid residue may be deleted. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

GBS 80

Examples of preferred GBS 80 fragments are discussed below. Polynucleotide and polypeptide sequences of GBS 80 from a variety of GBS serotypes and strain isolates are set forth in Figures 18 and 22. The polynucleotide and polypeptide sequences for GBS 80 from GBS serotype V, strain isolate 2603 are also included below as SEQ ID NOS 1 and 2:

SEQ ID NO. 1

ATGAAATTATCGAAGAAGTTATTGTTTTCGGCTGCTGTTTTAACAATGGTGGCGGGGTCAACTGTTGAACCAGTA
GCTCAGTTTGCGACTGGAATGAGTATTGTAAGAGCTGCAGAAGTGTCACAAGAACGCCCAGCGAAAACAACAGTA
AATATCTATAAATTACAAGCTGATAGTTATAAATCGGAAATTACTTCTAATGGTGGTATCGAGAATAAAGACGGC
GAAGTAATATCTAACTATGCTAAACTTGGTGACAATGTAAAAGGTTTGCAAGGTGTACAGTTTAAACGTTATAAA
GTCAAGACGGATATTTCTGTTGATGAATTGAAAAAATTGACAACAGTTGAAGCAGCAGATGCAAAAGTTGGAACG
ATTCTTGAAGAAGGTGTCAGTCTACCTCAAAAAACTAATGCTCAAGGTTTGGTCGTCGATGCTCTGGATTCAAAA
AGTAATGTGAGATACTTGTATGTAGAAGATTTAAAGAATTCACCTTCAAACATTACCAAAGCTTATGCTGTACCG
TTTGTGTTGGAATTACCAGTTGCTAACTCTACAGGTACAGGTTTCCTTTCTGAAATTAATATTTACCCTAAAAAC
GTTGTAACTGATGAACCAAAAACAGATAAAGATGTTAAAAAATTAGGTCAGGACGATGCAGGTTATACGATTGGT
GAAGAATTCAAATGGTTCTTGAAATCTACAATCCCTGCCAATTTAGGTGACTATGAAAAATTTGAAATTACTGAT
AAATTTGCAGATGGCTTGACTTATAAATCTGTTGGAAAAATCAAGATTGGTTCGAAAACACTGAATAGAGATGAG
CACTACACTATTGATGAACCAACAGTTGATAACCAAAATACATTAAAAATTACGTTTAAACCAGAGAAATTTAAA
GAAATTGCTGAGCTACTTAAAGGAATGACCCTTGTTAAAAATCAAGATGCTCTTGATAAAGCTACTGCAAATACA
GATGATGCGGCATTTTTGGAAATTCCAGTTGCATCAACTATTAATGAAAAAGCAGTTTTAGGAAAAGCAATTGAA
AATACTTTTGAACTTCAATATGACCATACTCCTGATAAAGCTGACAATCCAAAACCATCTAATCCTCCAAGAAAA
CCAGAAGTTCATACTGGTGGGAAACGATTTGTAAAGAAAGACTCAACAGAAACACAAACACTAGGTGGTGCTGAG
TTTGATTTGTTGGCTTCTGATGGGACAGCAGTAAAATGGACAGATGCTCTTATTAAAGCGAATACTAATAAAAAC
TATATTGCTGGAGAAGCTGTTACTGGGCAACCAATCAAATTGAAATCACATACAGACGGTACGTTTGAGATTAAA
GGTTTGGCTTATGCAGTTGATGCGAATGCAGAGGGTACAGCAGTAACTTACAAATTAAAAGAAACAAAAGCACCA
GAAGGTTATGTAATCCCTGATAAAGAAATCGAGTTTACAGTATCACAAACATCTTATAATACAAAACCAACTGAC
ATCACGGTTGATAGTGCTGATGCAACACCTGATACAATTAAAAACAACAAACGTCCTTCAATCCCTAATACTGGT
GGTATTGGTACGGCTATCTTTGTCGCTATCGGTGCTGCGGTGATGGCTTTTGCTGTTAAGGGGATGAAGCGTCGT
ACAAAAGATAAC

SEQ ID NO: 2

MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK

EEARLLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPS*IPNTG*
GIGTAIFVAIGAAVMAFAVKGMKRRTKDN

As described above, the compositions of the invention may include fragments of AI proteins. In some instances, removal of one or more domains, such as a leader or signal sequence region, a transmembrane region, a cytoplasmic region or a cell wall anchoring motif, may facilitate cloning of the gene encoding the protein and/or recombinant expression of the GBS AI protein. In addition, fragments comprising immunogenic epitopes of the cited GBS AI proteins may be used in the compositions of the invention.

For example, GBS 80 contains an N-terminal leader or signal sequence region which is indicated by the underlined sequence at the beginning of SEQ ID NO: 2 above. In one embodiment, one or more amino acids from the leader or signal sequence region of GBS 80 are removed. An example of such a GBS 80 fragment is set forth below as SEQ ID NO: 3:

SEQ ID NO: 3
AEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDGEVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKK
LTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSKSNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTG
TGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIGEEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVG
KIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFKEIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVAS
TINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRKPEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVK
WTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIKGLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEF
TVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPNTGGIGTAIFVAIGAAVMAFAVKGMKRRTKDN

GBS 80 contains a C-terminal transmembrane region which is indicated by the underlined sequence near the end of SEQ ID NO: 2 above. In one embodiment, one or more amino acids from the transmembrane region and/or a cytoplasmic region are removed. An example of such a GBS 80 fragment is set forth below as SEQ ID NO: 4:

SEQ ID NO: 4
MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK
EIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPS*IPNTG*

GBS 80 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 5** IPNTG (shown in italics in SEQ ID NO: 2 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 80 protein from the host cell. Accordingly, in one preferred fragment of GBS 80 for use in the invention, the transmembrane and/or cytoplasmic regions and the cell wall anchor motif are removed from GBS 80. An example of such a GBS 80 fragment is set forth below as SEQ ID NO: 6.

SEQ ID NO: 6
MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK

FTAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPS

Alternatively, in some recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

In one embodiment, the leader or signal sequence region, the transmembrane and cytoplasmic regions and the cell wall anchor motif are removed from the GBS 80 sequence. An example of such a GBS 80 fragment is set forth below as SEQ ID NO: 7.

SEQ ID NO: 7

AEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDGEVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKK
LTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSKSNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTG
TGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIGEEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVG
KIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFKEIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVAS
TINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRKPEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVK
WTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIKGLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEF
TVSQTSYNTKPTDITVDSADATPDTIKNNKRPS

Applicants have identified a particularly immunogenic fragment of the GBS 80 protein. This immunogenic fragment is located towards the N-terminus of the protein and is underlined in the GBS 80 SEQ ID NO: 2 sequence below. The underlined fragment is set forth below as SEQ ID NO: 8.

SEQ ID NO: 2

MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK
EIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPS*IPNTG*
GIGTAIFVAIGAAVMAFAVKGMKRRTKDN

SEQ ID NO: 8

AEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDGEVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKK
LTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSKSNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTG
TGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIGEEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVG
KIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFKEIAELLKG

The immunogenicity of the protein encoded by SEQ ID NO: 7 was compared against PBS, GBS whole cell, GBS 80 (full length) and another fragment of GBS 80, located closer to the C-terminus of the peptide (SEQ ID NO: 9, below).

SEQ ID NO: 9

MTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRKPEVHTGGK
RFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIKGLAYAVDA
NAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPS

Both an Active Maternal Immunization Assay and a Passive Maternal Immunization Assay were conducted on this collection of proteins.

As used herein, an Active Maternal Immunization assay refers to an *in vivo* protection assay where female mice are immunized with the test antigen composition. The female mice are then bred and their pups are challenged with a lethal dose of GBS. Serum titers of the female mice during the immunization schedule are measured as well as the survival time of the pups after challenge.

Specifically, the Active Maternal Immunization assays referred to herein used groups of four CD-1 female mice (Charles River Laboratories, Calco Italy). These mice were immunized intraperitoneally with the selected proteins in Freund's adjuvant at days 1, 21 and 35, prior to breeding. 6-8 weeks old mice received 20 μg protein/dose when immunized with a single antigen, 30-45 μg protein/dose (15 μg each antigen) when immunized with combination of antigens. The immune response of the dams was monitored by using serum samples taken on day 0 and 49. The female mice were bred 2-7 days after the last immunization (at approximately t= 36 – 37), and typically had a gestation period of 21 days. Within 48 hours of birth, the pups were challenged via I.P. with GBS in a dose approximately equal to a amount which would be sufficient to kill 70 – 90 % of unimmunized pups (as determined by empirical data gathered from PBS control groups). The GBS challenge dose is preferably administered in 50μl of THB medium. Preferably, the pup challenge takes place at 56 to 61 days after the first immunization. The challenge inocula were prepared starting from frozen cultures diluted to the appropriate concentration with THB prior to use. Survival of pups was monitored for 5 days after challenge.

As used herein, the Passive Maternal Immunization Assay refers to an *in vivo* protection assay where pregnant mice are passively immunized by injecting rabbit immune sera (or control sera) approximately 2 days before delivery. The pups are then challenged with a lethal dose of GBS.

Specifically, the Passive Maternal Immunization Assay referred to herein used groups of pregnant CD1 mice which were passively immunized by injecting 1 ml of rabbit immune sera or control sera via I.P., 2 days before delivery. Newborn mice (24-48 hrs after birth) are challenged via I.P. with a 70 - 90% lethal dose of GBS serotype III COH1. The challenge dose, obtained by diluting a frozen mid log phase culture, was administered in 50μl of THB medium.

For both assays, the number of pups surviving GBS infection was assessed every 12 hrs for 4 days. Statistical significance was estimated by Fisher's exact test.

The results of each assay for immunization with SEQ ID NO: 7, SEQ ID NO: 8, PBS and GBS whole cell are set forth in Tables 1 and 2 below.

TABLE 1: Immunization

Antigen	Alive/total	%Survival	Fisher's exact test
PBS (neg control)	13/80	16%	
GBS (whole cell)	54/65	83%	P<0.00000001
GBS80 (intact)	62/70	88%	P<0.00000001
GBS80 (fragment) SEQ ID 7	35/64	55%	P=0.0000013
GBS80 (fragment) SEQ ID 8	13/67	19%	P=0.66

Table 2: Passive Maternal Immunization

Antigen	Alive/total	%Survival	Fisher's exact test
PBS (neg control)	12/42	28%	
GBS (whole cell)	48/52	92%	P<0.00000001
GBS80 (intact)	48/55	87%	P<0.00000001
GBS80 (fragment) SEQ ID 7	45/57	79%	P=0.0000006
GBS80 (fragment) SEQ ID 8	13/54	24%	P=1

As shown in Tables 1 and 2, immunization with the SEQ ID NO: 7 GBS 80 fragment provided a substantially improved survival rate for the challenged pups than the comparison SEQ ID NO: 8 GBS 80 fragment. These results indicate that the SEQ ID NO: 7 GBS 80 fragment may comprise an important immunogenic epitope of GBS 80.

As discussed above, pilin motifs, containing conserved lysine (K) residues have been identified in GBS 80. The pilin motif sequences are underlined in SEQ ID NO: 2, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 199 and 207 and at amino acid residues 368 and 375. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of GBS 80. Preferred fragments of GBS 80 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 2

MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK
EIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPNTG
GIGTAIFVAIGAAVMAFAVKGMKRRTKDN

E boxes containing conserved glutamic residues have also been identified in GBS 80. The E box motifs are underlined in SEQ ID NO: 2 below. The conserved glutamic acid (E) residues, at amino acid residues 392 and 471, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of GBS 80. Preferred fragments of GBS 80 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 2

MKLSKKLLFSAAVLTMVAGSTVEPVAQFATGMSIVRAAEVSQERPAKTTVNIYKLQADSYKSEITSNGGIENKDG
EVISNYAKLGDNVKGLQGVQFKRYKVKTDISVDELKKLTTVEAADAKVGTILEEGVSLPQKTNAQGLVVDALDSK
SNVRYLYVEDLKNSPSNITKAYAVPFVLELPVANSTGTGFLSEINIYPKNVVTDEPKTDKDVKKLGQDDAGYTIG
EEFKWFLKSTIPANLGDYEKFEITDKFADGLTYKSVGKIKIGSKTLNRDEHYTIDEPTVDNQNTLKITFKPEKFK
EIAELLKGMTLVKNQDALDKATANTDDAAFLEIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPKPSNPPRK
PEVHTGGKRFVKKDSTETQTLGGAEFDLLASDGTAVKWTDALIKANTNKNYIAGEAVTGQPIKLKSHTDGTFEIK
GLAYAVDANAEGTAVTYKLKETKAPEGYVIPDKEIEFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPNTG
GIGTAIFVAIGAAVMAFAVKGMKRRTKDN

GBS 104

Similarly, the following offers examples of preferred GBS 104 fragments. Nucleotide and amino acid sequences of GBS 104 sequenced from serotype V isolated strain 2603 are set forth below as SEQ ID NOS 10 and 11:

SEQ ID NO. 10

ATGAAAAAGAGACAAAAAATATGGAGAGGGTTATCAGTTACTTTACTAATCCTGTCCCAAATTCCATTTGGTATA
TTGGTACAAGGTGAAACCCAAGATACCAATCAAGCACTTGGAAAAGTAATTGTTAAAAAAACGGGAGACAATGCT
ACACCATTAGGCAAAGCGACTTTTGTGTTAAAAAATGACAATGATAAGTCAGAAACAAGTCACGAAACGGTAGAG
GGTTCTGGAGAAGCAACCTTTGAAAACATAAAACCTGGAGACTACACATTAAGAGAAGAAACAGCACCAATTGGT
TATAAAAAAACTGATAAAACCTGGAAAGTTAAAGTTGCAGATAACGGAGCAACAATAATCGAGGGTATGGATGCA
GATAAAGCAGAGAAACGAAAAGAAGTTTTGAATGCCCAATATCCAAAATCAGCTATTTATGAGGATACAAAAGAA
AATTACCCATTAGTTAATGTAGAGGGTTCCAAAGTTGGTGAACAATACAAAGCATTGAATCCAATAAATGGAAAA
GATGGTCGAAGAGAGATTGCTGAAGGTTGGTTATCAAAAAAAATTACAGGGGTCAATGATCTCGATAAGAATAAA
TATAAAATTGAATTAACTGTTGAGGGTAAAACCACTGTTGAAACGAAAGAACTTAATCAACCACTAGATGTCGTT
GTGCTATTAGATAATTCAAATAGTATGAATAATGAAAGAGCCAATAATTCTCAAAGAGCATTAAAAGCTGGGGAA
GCAGTTGAAAAGCTGATTGATAAAATTACATCAAATAAAGACAATAGAGTAGCTCTTGTGACATATGCCTCAACC
ATTTTTGATGGTACTGAAGCGACCGTATCAAAGGGAGTTGCCGATCAAAATGGTAAAGCGCTGAATGATAGTGTA
TCATGGGATTATCATAAAACTACTTTTACAGCAACTACACATAATTACAGTTATTTAAATTTAACAAATGATGCT
AACGAAGTTAATATTCTAAAGTCAAGAATTCCAAAGGAAGCGGAGCATATAAATGGGGATCGCACGCTCTATCAA
TTTGGTGCGACATTTACTCAAAAAGCTCTAATGAAAGCAAATGAAATTTTAGAGACACAAAGTTCTAATGCTAGA
AAAAAACTTATTTTTCACGTAACTGATGGTGTCCCTACGATGTCTTATGCCATAAATTTTAATCCTTATATATCA
ACATCTTACCAAAACCAGTTTAATTCTTTTTTAAATAAAATACCAGATAGAAGTGGTATTCTCCAAGAGGATTTT
ATAATCAATGGTGATGATTATCAAATAGTAAAAGGAGATGGAGAGAGTTTTAAACTGTTTTCGGATAGAAAAGTT
CCTGTTACTGGAGGAACGACACAAGCAGCTTATCGAGTACCGCAAAATCAACTCTCTGTAATGAGTAATGAGGGA
TATGCAATTAATAGTGGATATATTTATCTCTATTGGAGAGATTACAACTGGGTCTATCCATTTGATCCTAAGACA
AAGAAAGTTTCTGCAACGAAACAAATCAAAACTCATGGTGAGCCAACAACATTATACTTTAATGGAAATATAAGA
CCTAAAGGTTATGACATTTTTACTGTTGGGATTGGTGTAAACGGAGATCCTGGTGCAACTCCTCTTGAAGCTGAG
AAATTTATGCAATCAATATCAAGTAAAACAGAAAATTATACTAATGTTGATGATACAAATAAAATTTATGATGAG
CTAAATAAATACTTTAAAACAATTGTTGAGGAAAAACATTCTATTGTTGATGGAAATGTGACTGATCCTATGGGA
GAGATGATTGAATTCCAATTAAAAAATGGTCAAAGTTTTACACATGATGATTACGTTTTGGTTGGAAATGATGGC
AGTCAATTAAAAAATGGTGTGGCTCTTGGTGGACCAAACAGTGATGGGGGAATTTTAAAAGATGTTACAGTGACT
TATGATAAGACATCTCAAACCATCAAAATCAATCATTTGAACTTAGGAAGTGGACAAAAAGTAGTTCTTACCTAT
GATGTACGTTTAAAAGATAACTATATAAGTAACAAATTTTACAATACAAATAATCGTACAACGCTAAGTCCGAAG
AGTGAAAAAGAACCAAATACTATTCGTGATTTCCCAATTCCCAAAATTCGTGATGTTCGTGAGTTTCCGGTACTA
ACCATCAGTAATCAGAAGAAAATGGGTGAGGTTGAATTTATTAAAGTTAATAAAGACAAACATTCAGAATCGCTT
TTGGGAGCTAAGTTTCAACTTCAGATAGAAAAAGATTTTTCTGGGTATAAGCAATTTGTTCCAGAGGGAAGTGAT
GTTACAACAAAGAATGATGGTAAAATTTATTTTAAAGCACTTCAAGATGGTAACTATAAATTATATGAAATTTCA
AGTCCAGATGGCTATATAGAGGTTAAAACGAAACCTGTTGTGACATTTACAATTCAAAATGGAGAAGTTACGAAC
CTGAAAGCAGATCCAAATGCTAATAAAAATCAAATCGGGTATCTTGAAGGAAATGGTAAACATCTTATTACCAAC
ACTCCCAAACGCCCACCAGGTGTTTTTCCTAAAACAGGGGGAATTGGTACAATTGTCTATATATTAGTTGGTTCT
ACTTTTATGATACTTACCATTTGTTCTTTCCGTCGTAAACAATTG

SEQ ID NO. 11

MKKRQKIWRGLSVTLLILSQIPFGILVQGETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVE
GSGEATFENIKPGDYTLREETAPIGYKKTDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKE
NYPLVNVEGSKVGEQYKALNPINGKDGRREIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVV
VLLDNSNSMNNERANNSQRALKAGEAVEKLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSV
SWDYHKTTFTATTHNYSYLNLTNDANEVNILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNAR
KKLIFHVTDGVPTMSYAINFNPYISTSYQNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV
PVTGGTTQAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIR
PKGYDIFTVGIGVNGDPGATPLEAEKFMQSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMG
EMIEFQLKNGQSFTHDDYVLVGNDGSQLKNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTY
DVRLKDNYISNKFYNTNNRTTLSPKSEKEPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESL
LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTN
LKADPNANKNQIGYLEGNGKHLITNTPKRPPGVFPKTGGIGTIVYILVGSTFMILTICSFRRKQL

GBS 104 contains an N-terminal leader or signal sequence region which is indicated by the underlined sequence at the beginning of SEQ ID NO 11 above. In one embodiment, one or more

amino acid sequences from the leader or signal sequence region of GBS 104 are removed. An example of such a GBS 104 fragment is set forth below as SEQ ID NO 12.

SEQ ID NO 12

GETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVEGSGEATFENIKPGDYTLREETAPIGYKK
TDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKENYPLVNVEGSKVGEQYKALNPINGKDGR
REIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVVVLLDNSNSMNNERANNSQRALKAGEAVE
KLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSVSWDYHKTTFTATTHNYSYLNLTNDANEV
NILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNARKKLIFHVTDGVPTMSYAINFNPYISTSY
QNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKVPVTGGTTQAAYRVPQNQLSVMSNEGYAI
NSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIRPKGYDIFTVGIGVNGDPGATPLEAEKFM
QSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMGEMIEFQLKNGQSFTHDDYVLVGNDGSQL
KNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTYDVRLKDNYISNKFYNTNNRTTLSPKSEK
EPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESLLGAKFQLQIEKDFSGYKQFVPEGSDVTT
KNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTNLKADPNANKNQIGYLEGNGKHLITNTPK
RPPGVFPKTGGIGTIVYILVGSTFMILTICSFRRKQL

GBS 104 contains a C-terminal transmembrane and/or cytoplasmic region which is indicated by the underlined region near the end of SEQ ID NO 11 above. In one embodiment, one or more amino acids from the transmembrane or cytomplasmic regions are removed. An example of such a GBS 104 fragment is set forth below as SEQ ID NO 13.

SEQ ID NO: 13

MKKRQKIWRGLSVTLLILSQIPFGILVQGETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVE
GSGEATFENIKPGDYTLREETAPIGYKKTDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKE
NYPLVNVEGSKVGEQYKALNPINGKDGRREIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVV
VLLDNSNSMNNERANNSQRALKAGEAVEKLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSV
SWDYHKTTFTATTHNYSYLNLTNDANEVNILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNAR
KKLIFHVTDGVPTMSYAINFNPYISTSYQNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV
PVTGGTTQAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIR
PKGYDIFTVGIGVNGDPGATPLEAEKFMQSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMG
EMIEFQLKNGQSFTHDDYVLVGNDGSQLKNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTY
DVRLKDNYISNKFYNTNNRTTLSPKSEKEPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESL
LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTN
LKADPNANKNQIGYLEGNGKHLITNT

In one embodiment, one or more amino acids from the leader or signal sequence region and one or more amino acids from the transmembrane or cytoplasmic regions are removed. An example of such a GBS 104 fragment is set forth below as SEQ ID NO 14.

SEQ ID NO: 14

GETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVEGSGEATFENIKPGDYTLREETAPIGYKK
TDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKENYPLVNVEGSKVGEQYKALNPINGKDGR
REIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVVVLLDNSNSMNNERANNSQRALKAGEAVE
KLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSVSWDYHKTTFTATTHNYSYLNLTNDANEV
NILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNARKKLIFHVTDGVPTMSYAINFNPYISTSY
QNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKVPVTGGTTQAAYRVPQNQLSVMSNEGYAI
NSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIRPKGYDIFTVGIGVNGDPGATPLEAEKFM
QSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMGEMIEFQLKNGQSFTHDDYVLVGNDGSQL
KNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTYDVRLKDNYISNKFYNTNNRTTLSPKSEK
EPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESLLGAKFQLQIEKDFSGYKQFVPEGSDVTT
KNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTNLKADPNANKNQIGYLEGNGKHLITNT

GBS 104, like GBS 80, contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 123** FPKTG (shown in italics in SEQ ID NO: 11 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 104 protein from the host cell. Accordingly, in one preferred fragment of GBS 104 for use in the

invention, only the transmembrane and/or cytoplasmic regions and the cell wall anchor motif are removed from GBS 104. Alternatively, in some recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, containing conserved lysine (K) residues, have been identified in GBS 104. The pilin motif sequences are underlined in SEQ ID NO: 11, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 141 and 149 and at amino acid residues 499 and 507. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of GBS 104. Preferred fragments of GBS 104 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO. 11

MKKRQKIWRGLSVTLLILSQIPFGILVQGETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVE
GSGEATFENIKPGDYTLREETAPIGYKKTDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKE
NYPLVNVEGSKVGEQYKALNPINGKDGRREIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVV
VLLDNSNSMNNERANNSQRALKAGEAVEKLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSV
SWDYHKTTFTATTHNYSYLNLTNDANEVNILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNAR
KKLIFHVTDGVPTMSYAINFNPYISTSYQNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV
PVTGGTTQAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIR
PKGYDIFTVGIGVNGDPGATPLEAEKFMQSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMG
EMIEFQLKNGQSFTHDDYVLVGNDGSQLKNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTY
DVRLKDNYISNKFYNTNNRTTLSPKSEKEPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESL
LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTN
LKADPNANKNQIGYLEGNGKHLITNTPKRPPGVFPKTGGIGTIVYILVGSTFMILTICSFRRKQL

Two E boxes containing a conserved glutamic residues have also been identified in GBS 104. The E box motifs are underlined in SEQ ID NO: 11 below. The conserved glutamic acid (E) residues, at amino acid residues 94 and 798, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of GBS 104. Preferred fragments of GBS 104 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO. 11

MKKRQKIWRGLSVTLLILSQIPFGILVQGETQDTNQALGKVIVKKTGDNATPLGKATFVLKNDNDKSETSHETVE
GSGEATFENIKPGDYTLREETAPIGYKKTDKTWKVKVADNGATIIEGMDADKAEKRKEVLNAQYPKSAIYEDTKE
NYPLVNVEGSKVGEQYKALNPINGKDGRREIAEGWLSKKITGVNDLDKNKYKIELTVEGKTTVETKELNQPLDVV
VLLDNSNSMNNERANNSQRALKAGEAVEKLIDKITSNKDNRVALVTYASTIFDGTEATVSKGVADQNGKALNDSV
SWDYHKTTFTATTHNYSYLNLTNDANEVNILKSRIPKEAEHINGDRTLYQFGATFTQKALMKANEILETQSSNAR
KKLIFHVTDGVPTMSYAINFNPYISTSYQNQFNSFLNKIPDRSGILQEDFIINGDDYQIVKGDGESFKLFSDRKV
PVTGGTTQAAYRVPQNQLSVMSNEGYAINSGYIYLYWRDYNWVYPFDPKTKKVSATKQIKTHGEPTTLYFNGNIR
PKGYDIFTVGIGVNGDPGATPLEAEKFMQSISSKTENYTNVDDTNKIYDELNKYFKTIVEEKHSIVDGNVTDPMG
EMIEFQLKNGQSFTHDDYVLVGNDGSQLKNGVALGGPNSDGGILKDVTVTYDKTSQTIKINHLNLGSGQKVVLTY
DVRLKDNYISNKFYNTNNRTTLSPKSEKEPNTIRDFPIPKIRDVREFPVLTISNQKKMGEVEFIKVNKDKHSESL
LGAKFQLQIEKDFSGYKQFVPEGSDVTTKNDGKIYFKALQDGNYKLYEISSPDGYIEVKTKPVVTFTIQNGEVTN
LKADPNANKNQIGYLEGNGKHLITNTPKRPPGVFPKTGGIGTIVYILVGSTFMILTICSFRRKQL

GBS 067

The following offers examples of preferred GBS 067 fragments. Nucleotide and amino acid sequence of GBS 067 sequences from serotype V isolated strain 2603 are set forth below as SEQ ID NOS: 15 and 16.

PCT/US05/27239

SEQ ID NO: 15

ATGAGAAAATACCAAAAATTTTCTAAAATATTGACGTTAAGTCTTTTTTGTTTGTCGCAAATACCGCTTAATACC
AATGTTTTAGGGGAAAGTACCGTACCGGAAAATGGTGCTAAAGGAAAGTTAGTTGTTAAAAAGACAGATGACCAG
AACAAACCACTTTCAAAAGCTACCTTTGTTTTAAAAACTACTGCTCATCCAGAAAGTAAAATAGAAAAAGTAACT
GCTGAGCTAACAGGTGAAGCTACTTTTGATAATCTCATACCTGGAGATTATACTTTATCAGAAGAAACAGCGCCC
GAAGGTTATAAAAAGACTAACCAGACTTGGCAAGTTAAGGTTGAGAGTAATGGAAAAACTACGATACAAAATAGT
GGTGATAAAAATTCCACAATTGGACAAAATCAGGAAGAACTAGATAAGCAGTATCCCCCCACAGGAATTTATGAA
GATACAAAGGAATCTTATAAACTTGAGCATGTTAAAGGTTCAGTTCCAAATGGAAAGTCAGAGGCAAAAGCAGTT
AACCCATATTCAAGTGAAGGTGAGCATATAAGAGAAATTCCAGAGGGAACATTATCTAAACGTATTTCAGAAGTA
GGTGATTTAGCTCATAATAAATATAAAATTGAGTTAACTGTCAGTGGAAAAACCATAGTAAAACCAGTGGACAAA
CAAAAGCCGTTAGATGTTGTCTTCGTACTCGATAATTCTAACTCAATGAATAACGATGGCCCAAATTTTCAAAGG
CATAATAAAGCCAAGAAAGCTGCCGAAGCTCTTGGGACCGCAGTAAAAGATATTTTAGGAGCAAACAGTGATAAT
AGGGTTGCATTAGTTACCTATGGTTCAGATATTTTTGATGGTAGGAGTGTAGATGTCGTAAAAGGATTTAAAGAA
GATGATAAATATTATGGCCTTCAAACTAAGTTCACAATTCAGACAGAGAATTATAGTCATAAACAATTAACAAAT
AATGCTGAAGAGATTATAAAAAGGATTCCGACAGAAGCTCCTAAAGCTAAGTGGGGATCTACTACCAATGGATTA
ACTCCAGAGCAACAAAAGGAGTACTATCTTAGTAAAGTAGGAGAAACATTTACTATGAAAGCCTTCATGGAGGCA
GATGATATTTTGAGTCAAGTAAATCGAAATAGTCAAAAAATTATTGTTCATGTAACTGATGGTGTTCCTACGAGA
TCATATGCTATTAATAATTTTAAACTGGGTGCATCATATGAAAGCCAATTTGAACAAATGAAAAAAAATGGATAT
CTAAATAAAAGTAATTTTCTACTTACTGATAAGCCCGAGGATATAAAAGGAAATGGGGAGAGTTACTTTTTGTTT
CCCTTAGATAGTTATCAAACACAGATAATCTCTGGAAACTTACAAAAACTTCATTATTTAGATTTAAATCTTAAT
TACCCTAAAGGTACAATTTATCGAAATGGACCAGTGAAAGAACATGGAACACCAACCAAACTTTATATAAATAGT
TTAAAACAGAAAAATTATGACATTTTTAATTTTGGTATCGATATATCTGGTTTTAGACAAGTTTATAATGAGGAG
TATAAGAAAAATCAAGATGGTACTTTTCAAAAATTGAAAGAGGAAGCTTTTAAACTTTCAGATGGAGAAATCACA
GAACTAATGAGGTCGTTCTCTTCCAAACCTGAGTACTACACCCCTATCGTAACTTCAGCCGATACATCTAACAAT
GAAATTTTATCTAAAATTCAGCAACAATTTGAAACGATTTTAACAAAAGAAAACTCAATTGTTAATGGAACTATC
GAAGATCCTATGGGTGATAAAATCAATTTACAGCTTGGTAATGGACAAACATTACAGCCAAGTGATTATACTTTA
CAGGGAAATGATGGAAGTGTAATGAAGGATGGTATTGCAACTGGTGGGCCTAATAATGATGGTGGAATACTTAAG
GGGGTTAAATTAGAATACATCGGAAATAAACTCTATGTTAGAGGTTTGAATTTAGGAGAAGGTCAAAAAGTAACA
CTCACATATGATGTGAAACTAGATGACAGTTTTATAAGTAACAAATTCTATGACACTAATGGTAGAACAACATTG
AATCCTAAGTCAGAGGATCCTAATACACTTAGAGATTTTCCAATCCCTAAAATTCGTGATGTGAGAGAATATCCT
ACAATAACGATTAAAAACGAGAAGAAGTTAGGTGAAATTGAATTTATAAAAGTTGATAAAGATAATAATAAGTTG
CTTCTCAAAGGAGCTACGTTTGAACTTCAAGAATTTAATGAAGATTATAAACTTTATTTACCAATAAAAAATAAT
AATTCAAAAGTAGTGACGGGAGAAAACGGCAAAATTTCTTACAAAGATTTGAAAGATGGCAAATATCAGTTAATA
GAAGCAGTTTCGCCGGAGGATTATCAAAAAATTACTAATAAACCAATTTTAACTTTTGAAGTGGTTAAAGGATCG
ATAAAAAATATAATAGCTGTTAATAAACAGATTTCTGAATATCATGAGGAAGGTGACAAGCATTTAATTACCAAC
ACGCATATTCCACCAAAAGGAATTATTCCTATGACAGGTGGGAAAGGAATTCTATCTTTCATTTTAATAGGTGGA
GCTATGATGTCTATTGCAGGTGGAATTTATATTTGGAAAAGGTATAAGAAATCTAGTGATATGTCCATCAAAAAA
GAT

SEQ ID NO: 16

MRKYQKFSKILTLSLFCLSQIPLNTNVLGESTVPENGAKGKLVVKKTDDQNKPLSKATFVLKTTAHPESKIEKVT
AELTGEATFDNLIPGDYTLSEETAPEGYKKTNQTWQVKVESNGKTTIQNSGDKNSTIGQNQEELDKQYPPTGIYE
DTKESYKLEHVKGSVPNGKSEAKAVNPYSSEGEHIREIPEGTLSKRISEVGDLAHNKYKIELTVSGKTIVKPVDK
QKPLDVVFVLDNSNSMNNDGPNFQRHNKAKKAAEALGTAVKDILGANSDNRVALVTYGSDIFDGRSVDVVKGFKE
DDKYYGLQTKFTIQTENYSHKQLTNNAEEIIKRIPTEAPKAKWGSTTNGLTPEQQKEYYLSKVGETFTMKAFMEA
DDILSQVNRNSQKIIVHVTDGVPTRSYAINNFKLGASYESQFEQMKKNGYLNKSNFLLTDKPEDIKGNGESYFLF
PLDSYQTQIISGNLQKLHYLDLNLNYPKGTIYRNGPVKEHGTPTKLYINSLKQKNYDIFNFGIDISGFRQVYNEE
YKKNQDGTFQKLKEEAFKLSDGEITELMRSFSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKENSIVNGTI
EDPMGDKINLQLGNGQTLQPSDYTLQGNDGSVMKDGIATGGPNNDGGILKGVKLEYIGNKLYVRGLNLGEGQKVT
LTYDVKLDDSFISNKFYDTNGRTTLNPKSEDPNTLRDFPIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNNKL
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLIEAVSPEDYQKITNKPILTFEVVKGS
IKNIIAVNKQISEYHEEGDKHLITNTHIPPKGI*IPMTGG*KGILSFILIGGAMMSIAGGIYIWKRYKKSSDMSIKK
D

GBS 067 contains a C-terminus transmembrane region which is indicated by the underlined region closest to the C-terminus of SEQ ID NO: 16 above. In one embodiment, one or more amino acids from the transmembrane region is removed and or the amino acid is truncated before the

transmembrane region. An example of such a GBS 067 fragment is set forth below as SEQ ID NO: 17.

SEQ ID NO: 17

MRKYQKFSKILTLSLFCLSQIPLNTNVLGESTVPENGAKGKLVVKKTDDQNKPLSKATFVLKTTAHPESKIEKVT
AELTGEATFDNLIPGDYTLSEETAPEGYKKTNQTWQVKVESNGKTTIQNSGDKNSTIGQNQEELDKQYPPTGIYE
DTKESYKLEHVKGSVPNGKSEAKAVNPYSSEGEHIREIPEGTLSKRISEVGDLAHNKYKIELTVSGKTIVKPVDK
QKPLDVVFVLDNSNSMNNDGPNFQRHNKAKKAAEALGTAVKDILGANSDNRVALVTYGSDIFDGRSVDVVKGFKE
DDKYYGLQTKFTIQTENYSHKQLTNNAEEIIKRIPTEAPKAKWGSTTNGLTPEQQKEYYLSKVGETFTMKAFMEA
DDILSQVNRNSQKIIVHVTDGVPTRSYAINNFKLGASYESQFEQMKKNGYLNKSNFLLTDKPEDIKGNGESYFLF
PLDSYQTQIISGNLQKLHYLDLNLNYPKGTIYRNGPVKEHGTPTKLYINSLKQKNYDIFNFGIDISGFRQVYNEE
YKKNQDGTFQKLKEEAFKLSDGEITELMRSFSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKENSIVNGTI
EDPMGDKINLQLGNGQTLQPSDYTLQGNDGSVMKDGIATGGPNNDGGILKGVKLEYIGNKLYVRGLNLGEGQKVT
LTYDVKLDDSFISNKFYDTNGRTTLNPKSEDPNTLRDFPIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNNKL
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLIEAVSPEDYQKITNKPILTFEVVKGS
IKNIIAVNKQISEYHEEGDKHLITNTHIPPKGIIPMTGGKGILS

GBS 067 contains an amino acid motif indicative of a cell wall anchor (an LPXTG (SEQ ID NO: 122) motif): **SEQ ID NO: 18** IPMTG. (shown in italics in SEQ ID NO: 16 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 067 protein from the host cell. Accordingly, in one preferred fragment of GBS 067 for use in the invention, the transmembrane and the cell wall anchor motif are removed from GBS 67. An example of such a GBS 067 fragment is set forth below as SEQ ID NO: 19.

SEQ ID NO: 19

MRKYQKFSKILTLSLFCLSQIPLNTNVLGESTVPENGAKGKLVVKKTDDQNKPLSKATFVLKTTAHPESKIEKVT
AELTGEATFDNLIPGDYTLSEETAPEGYKKTNQTWQVKVESNGKTTIQNSGDKNSTIGQNQEELDKQYPPTGIYE
DTKESYKLEHVKGSVPNGKSEAKAVNPYSSEGEHIREIPEGTLSKRISEVGDLAHNKYKIELTVSGKTIVKPVDK
QKPLDVVFVLDNSNSMNNDGPNFQRHNKAKKAAEALGTAVKDILGANSDNRVALVTYGSDIFDGRSVDVVKGFKE
DDKYYGLQTKFTIQTENYSHKQLTNNAEEIIKRIPTEAPKAKWGSTTNGLTPEQQKEYYLSKVGETFTMKAFMEA
DDILSQVNRNSQKIIVHVTDGVPTRSYAINNFKLGASYESQFEQMKKNGYLNKSNFLLTDKPEDIKGNGESYFLF
PLDSYQTQIISGNLQKLHYLDLNLNYPKGTIYRNGPVKEHGTPTKLYINSLKQKNYDIFNFGIDISGFRQVYNEE
YKKNQDGTFQKLKEEAFKLSDGEITELMRSFSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKENSIVNGTI
EDPMGDKINLQLGNGQTLQPSDYTLQGNDGSVMKDGIATGGPNNDGGILKGVKLEYIGNKLYVRGLNLGEGQKVT
LTYDVKLDDSFISNKFYDTNGRTTLNPKSEDPNTLRDFPIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNNKL
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLIEAVSPEDYQKITNKPILTFEVVKGS
IKNIIAVNKQISEYHEEGDKHLITNTHIPPKGI

Alternatively, in some recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Three pilin motifs, containing conserved lysine (K) residues have been identified in GBS 67. The pilin motif sequences are underlined in SEQ ID NO: 16, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 478 and 488, at amino acid residues 340 and 342, and at amino acid residues 703 and 717. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of GBS 67. Preferred fragments of GBS 67 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 16

MRKYQKFSKILTLSLFCLSQIPLNTNVLGESTVPENGAKGKLVVKKTDDQNKPLSKATFVLKTTAHPESKIEKVT
AELTGEATFDNLIPGDYTLSEETAPEGYKKTNQTWQVKVESNGKTTIQNSGDKNSTIGQNQEELDKQYPPTGIYE
DTKESYKLEHVKGSVPNGKSEAKAVNPYSSEGEHIREIPEGTLSKRISEVGDLAHNKYKIELTVSGKTIVKPVDK
QKPLDVVFVLDNSNSMNNDGPNFQRHNKAKKAAEALGTAVKDILGANSDNRVALVTYGSDIFDGRSVDVVKGFKE
DDKYYGLQTKFTIQTENYSHKQLTNNAEEIIKRIPTEAPKAKWGSTTNGLTPEQQKEYYLSKVGETFTMKAFMEA
DDILSQVNRNSQKIIVHVTDGVPTRSYAINNFKLGASYESQFEQMKKNGYLNKSNFLLTDKPEDIKGNGESYFLF
PLDSYQTQIISGNLQKLHYLDLNLNYPKGTIYRNGPVKEHGTPTKLYINSLKQKNYDIFNFGIDISGFRQVYNEE
YKKNQDGTFQKLKEEAFKLSDGEITELMRSFSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKENSIVNGTI
EDPMGDKINLQLGNGQTLQPSDYTLQGNDGSVMKDGIATGGPNNDGGILKGVKLEYIGNKLYVRGLNLGEGQKVT
LTYDVKLDDSFISNKFYDTNGRTTLNPKSEDPNTLRDFPIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNNKL
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLIEAVSPEDYQKITNKPILTFEVVKGS
IKNIIAVNKQISEYHEEGDKHLITNTHIPPKGIIPMTGGKGILSFILIGGAMMSIAGGIYIWKRYKKSSDMSIKK
D

Two E boxes containing conserved glutamic residues have also been identified in GBS 67. The E box motifs are underlined in SEQ ID NO: 16 below. The conserved glutamic acid (E) residues, at amino acid residues 96 and 801, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of GBS 67. Preferred fragments of GBS 67 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 16

MRKYQKFSKILTLSLFCLSQIPLNTNVLGESTVPENGAKGKLVVKKTDDQNKPLSKATFVLKTTAHPESKIEKVT
AELTGEATFDNLIPGDYTLSEETAPEGYKKTNQTWQVKVESNGKTTIQNSGDKNSTIGQNQEELDKQYPPTGIYE
DTKESYKLEHVKGSVPNGKSEAKAVNPYSSEGEHIREIPEGTLSKRISEVGDLAHNKYKIELTVSGKTIVKPVDK
QKPLDVVFVLDNSNSMNNDGPNFQRHNKAKKAAEALGTAVKDILGANSDNRVALVTYGSDIFDGRSVDVVKGFKE
DDKYYGLQTKFTIQTENYSHKQLTNNAEEIIKRIPTEAPKAKWGSTTNGLTPEQQKEYYLSKVGETFTMKAFMEA
DDILSQVNRNSQKIIVHVTDGVPTRSYAINNFKLGASYESQFEQMKKNGYLNKSNFLLTDKPEDIKGNGESYFLF
PLDSYQTQIISGNLQKLHYLDLNLNYPKGTIYRNGPVKEHGTPTKLYINSLKQKNYDIFNFGIDISGFRQVYNEE
YKKNQDGTFQKLKEEAFKLSDGEITELMRSFSSKPEYYTPIVTSADTSNNEILSKIQQQFETILTKENSIVNGTI
EDPMGDKINLQLGNGQTLQPSDYTLQGNDGSVMKDGIATGGPNNDGGILKGVKLEYIGNKLYVRGLNLGEGQKVT
LTYDVKLDDSFISNKFYDTNGRTTLNPKSEDPNTLRDFPIPKIRDVREYPTITIKNEKKLGEIEFIKVDKDNNKL
LLKGATFELQEFNEDYKLYLPIKNNNSKVVTGENGKISYKDLKDGKYQLIEAVSPEDYQKITNKPILTFEVVKGS
IKNIIAVNKQISEYHEEGDKHLITNTHIPPKGIIPMTGGKGILSFILIGGAMMSIAGGIYIWKRYKKSSDMSIKK
D

Predicted secondary structure for the GBS 067 amino acid sequence is set forth in FIGURE 33. As shown in this figure, GBS 067 contains several regions predicted to form alpha helical structures. Such alpha helical regions are likely to form coiled-coil structures and may be involved in oligomerization of GBS 067.

The amino acid sequence for GBS 067 also contains a region which is homologous to the Cna_B domain of the Staphylococcus aureus collagen-binding surface protein (pfam05738). Although the Cna_B region is not thought to mediate collagen binding, it is predicted to form a beta sandwich structure. In the Staph aureus protein, this beta sandwich structure is through to form a stalk that presents the ligand binding domain away from the bacterial cell surface. This same amino acid sequence region is also predicted to be an outer membrane protein involved in cell envelope biogenesis.

The amino acid sequence for GBS 067 contains a region which is homologous to a von Willebrand factor (vWF) type A domain. The vWF type A domain is present at amino acid residues 229-402 of GBS 067 as shown in SEQ ID NO: 16. This type of sequence is typically found in

extracellular proteins such as integrins and it thought to mediate adhesion, including adhesion to collagen, fibronectin, and fibrinogen, discussed above.

Because applicants have identified GBS 67 as a surface exposed protein on GBS and because GBS 67 may be involved in GBS adhesion, the immunogenicity of the GBS 67 protein was examined in mice. The results of an immunization assay with GBS 67 are set forth in Table 48, below.

Table 48: GBS 67 Protects Mice in an Immunization Assay

Challenge GBS strain (serotype)	GBS 67 immungen		PBS immunogen		FACS
	dead/treated	% survival	dead/treated	% survival	Δmean
3050 (II)	0/30	100	29/49	41	460
CJB111 (V)	76/185	59	143/189	24	481
7357 b (Ib)	34/56	39	65/74	12	316

As shown in Table 48, immunization with GBS 67 provides a substantially improved survival rate for challenged mice relative to negative control, PBS, immunized mice. These results indicate that GBS 67 may comprise an immunogenic composition of the invention.

GBS 59

The following offers examples of GBS 59 fragments. Nucleotide and amino acid sequences of GBS 59 sequenced from serotype V isolated strain 2603 are set forth below as SEQ ID NOS: 125 and 126. The GBS 59 polypeptide of SEQ ID NO: 126 is referred to as SAG1407.

SEQ ID NO: 125

ttaagcttcctttgattggcgtcttttcatgataactactgctccaagcataatgcttaaaccaataattgtgaa
aagaattgtaccaataccacctgtttgtgggattgttacctttttatttttctacacgtgtcgcatctttttggtt
gctgttagcaacgtagtcaatgttaccacctgttatgtatgacccttgattaactacaaacttaatattacctgc
caacttagcaaatcctgctggagcaagtgtttcttcaaggttgtaagtaccgtctgcaagacctgtaacttcaaa
ttgaccttgatcgtttgaagtgtaggtaatggctctagccttatctgttatccactcataagctgtacgagcctc
aatgaaggctgcatcgtaatctgcttgtttagttttgataagttctttttgcagtaattccttttttcacctttttg
gtctgttgcagacaacttgttataagcagcgatagcttcatctaaagctatttttcttagcagctaaagttttttg
accttctgattgatctgctttaagagcaaggtatttacctgctgagtttttcacaacgaattgtgcaccagccaa
acggtcaccttgttcattagttttgacaaatttcttaccatgagtttcaacttttggttcagttgggttcaatgg
tgttgggttatcagaatctttggtattggtaatggttactttaccatttttctagatttattgcacttccgtaacc
agaaacacgttctgagatcatgtatgatttgttttctagaccagtgaatttacccgagaagttaccagatacttc
aaatttgataccatttccaaggtcgattgtacctttagatgtttttgtcaatgatactgaagcaacagttttatc
tttatctttcaatgtgtaaacaacgtttacaccatcaggtgcaattccgtcagaccaagttttagcaactgttac
ttcacccttgaaggtgtaacaggaagttcagtcaagtctttacctggtttgttaccatacgacaatttgatatc
attggattctggattatcaataattgcttgaccattaacagtagcactataagtcaatgtaaattcaatatcagc
tgttttagctgctttttccaatttgcccaatccatcagctgtgaattttaatgtgaaaccacgggcatcaatgct
aagttcatagtctgtatccttagcaaaagtttctgtagttcctgaagctttaaggctaacagttgaacccattgt
caaaccatttgacattatatctgtccaaaccaagttttcgtatttagaacctttgtgaatttttgttttaacttc
ataaggaacaactttaccgatttcagcagtagcagttgctttgtcacgtgcataattaccataatttgcgccagc
tgtcaaaagtctattaacatctgtcaatgctgtcaaatcgtttgttttagcaaagtttttatcaatttctggttt
ttcttcagtgttctttggataaacatgggcatcagcaacaacaccatcttcatttaccaatggaagagtgatgtt
aactggaaccgcttttgaagcagccaggagggaaccattattgttgtaagtagattttgatttaacttcaacaat
tttaaactcgcctttcaatcctttggtgttgaaaacaagtccagtatctccctctggtgtcaatccagacacggc
ctcatcaatatttactgttatttcaggagtaccatctttattaattaaggctggtgttaatttgttaccttcttt
tgccttaacatattgcactttaccacttttatcttctttcaaagctaaagcaaagaacgcaccttcgatttcttt
agatccctcgccaaagtaaccagcaaggtcagaaatagctccacctttgtagtcttttccgttaagacctgtagt
tcctgggaagttacttttgttaagatttgattcggtttgcaaaatcttgtgcaaagtcactgtattagttgttgc

[illegible]aatgacgttaaagtcagtaacaatgccgagaacattgcaaaata
tttgttgattcttttcat

SEQ ID NO: 126

MKRINKYFAMFSALLLTLTSLLSVAPAFADEATTNTVTLHKILQTESNLNKSNFPGTTGLNGKDYKGGAISDLAG
YFGEGSKEIEGAFFALALKEDKSGKVQYVKAKEGNKLTPALINKDGTPEITVNIDEAVSGLTPEGDTGLVFNTKG
LKGEFKIVEVKSKSTYNNNGSLLAASKAVPVNITLPLVNEDGVVADAHVYPKNTEEKPEIDKNFAKTNDLTALTD
VNRLLTAGANYGNYARDKATATAEIGKVVPYEVKTKIHKGSKYENLVWTDIMSNGLTMGSTVSLKASGTTETFAK
DTDYELSIDARGFTLKFTADGLGKLEKAAKTADIEFTLTYSATVNGQAIIDNPESNDIKLSYGNKPGKDLTELPV
TPSKGEVTVAKTWSDGIAPDGVNVVYTLKDKDKTVASVSLTKTSKGTIDLGNGIKFEVSGNFSGKFTGLENKSYM
ISERVSGYGSAINLENGKVTITNTKDSDNPTPLNPTEPKVETHGKKFVKTNEQGDRLAGAQFVVKNSAGKYLALK
ADQSEGQKTLAAKKIALDEAIAAYNKLSATDQKGEKGITAKELIKTKQADYDAAFIEARTAYEWITDKARAITYT
SNDQGQFEVTGLADGTYNLEETLAPAGFAKLAGNIKFVVNQGSYITGGNIDYVANSNQKDATRVENKKVT*IPQTG*
GIGTILFTIIGLSIMLGAVVIMKRRQSKEA

Nucleotide and amino acid sequences of GBS 59 sequenced from serotype V isolated strain CJB111 are set forth below as SEQ ID NOS: 127 and 128. The GBS 59 polypeptide of SEQ ID NO: 128 is referred to as BO1575.

SEQ ID NO: 127

ATGAAAAAAATCAACAAATGTCTTACAATGTTCTCGACACTGCTATTGATCTTAACGTCACTATTCTCAGTTGCA
CCAGCGTTTGCGGACGACGCAACAACTGATACTGTGACCTTGCACAAGATTGTCATGCCACAAGCTGCATTTGAT
AACTTTACTGAAGGTACAAAAGGTAAGAATGATAGCGATTATGTTGGTAAACAAATTAATGACCTTAAATCTTAT
TTTGGCTCAACCGATGCTAAAGAAATCAAGGGTGCTTTCTTTGTTTTCAAAAATGAAACTGGTACAAAATTCATT
ACTGAAAATGGTAAGGAAGTCGATACTTTGGAAGCTAAAGATGCTGAAGGTGGTGCTGTTCTTTCAGGGTTAACA
AAAGACAATGGTTTTGTTTTTAACACTGCTAAGTTAAAAGGAATTTACCAAATCGTTGAATTGAAAGAAAAATCA
AACTACGATAACAACGGTTCTATCTTGGCTGATTCAAAAGCAGTTCCAGTTAAAATCACTCTGCCATTGGTAAAC
AACCAAGGTGTTGTTAAAGATGCTCACATTTATCCAAAGAATACTGAAACAAAACCACAAGTAGATAAGAACTTT
GCAGATAAAGATCTTGATTATACTGACAACCGAAAAGACAAAGGTGTTGTCTCAGCGACAGTTGGTGACAAAAAA
GAATACATAGTTGGAACAAAAATTCTTAAAGGCTCAGACTATAAGAAACTGGTTTGGACTGATAGCATGACTAAA
GGTTTGACGTTCAACAACAACGTTAAAGTAACATTGGATGGTGAAGATTTTCCTGTTTTAAACTACAAACTCGTA
ACAGATGACCAAGGTTTCCGTCTTGCCTTGAATGCAACAGGTCTTGCAGCAGTAGCAGCAGCTGCAAAAGACAAA
GATGTTGAAATCAAGATCACTTACTCAGCTACGGTGAACGGCTCCACTACTGTTGAAATTCCAGAAACCAATGAT
GTTAAATTGGACTATGGTAATAACCCAACGGAAGAAAGTGAACCACAAGAAGGTACTCCAGCTAACCAAGAAATT
AAAGTCATTAAAGACTGGGCAGTAGATGGTACAATTACTGATGCTAATGTTGCAGTTAAAGCTATCTTTACCTTG
CAAGAAAAACAAACGGATGGTACATGGGTGAACGTTGCTTCACACGAAGCAACAAAACCATCACGCTTTGAACAT
ACTTTCACAGGTTTGGATAATGCTAAAACTTACCGCGTTGTCGAACGTGTTAGCGGCTACACTCCAGAATACGTA
TCATTTAAAAATGGTGTTGTGACTATCAAGAACAACAAAAACTCAAATGATCCAACTCCAATCAACCCATCAGAA
CCAAAAGTGGTGACTTATGGACGTAAATTTGTGAAAACAAATCAAGCTAACACTGAACGCTTGGCAGGAGCTACC
TTCCTCGTTAAGAAAGAAGGCAAATACTTGGCACGTAAAGCAGGTGCAGCAACTGCTGAAGCAAAGGCAGCTGTA
AAAACTGCTAAACTAGCATTGGATGAAGCTGTTAAAGCTTATAACGACTTGACTAAAGAAAAACAAGAAGGCCAA
GAAGGTAAAACAGCATTGGCTACTGTTGATCAAAAACAAAAAGCTTACAATGACGCTTTTGTTAAAGCTAACTAC
TCATATGAATGGGTTGCAGATAAAAAGGCTGATAATGTTGTTAAATTGATCTCTAACGCCGGTGGTCAATTTGAA
ATTACTGGTTTGGATAAAGGCACTTATGGCTTGGAAGAAACTCAAGCACCAGCAGGTTATGCGACATTGTCAGGT
GATGTAAACTTTGAAGTAACTGCCACATCATATAGCAAAGGGGCTACAACTGACATCGCATATGATAAAGGCTCT
GTAAAAAAAGATGCCCAACAAGTTCAAAACAAAAAAGTAACCATCCCACAAACAGGTGGTATTGGTACAATTCTT
TTCACAATTATTGGTTTAAGCATTATGCTTGGAGCAGTAGTTATCATGAAAAAACGTCAATCAGAGGAAGCTTAA

SEQ ID NO: 128

MKKINKCLTMFSTLLLILTSLFSVAPAFADDATTDTVTLHKIVMPQAAFDNFTEGTKGKNDSDYVGKQINDLKSY
FGSTDAKEIKGAFFVFKNETGTKFITENGKEVDTLEAKDAEGGAVLSGLTKDNGFVFNTAKLKGIYQIVELKEKS
NYDNNGSILADSKAVPVKITLPLVNNQGVVKDAHIYPKNTETKPQVDKNFADKDLDYTDNRKDKGVVSATVGDKK
EYIVGTKILKGSDYKKLVWTDSMTKGLTFNNNVKVTLDGEDFPVLNYKLVTDDQGFRLALNATGLAAVAAAAKDK
DVEIKITYSATVNGSTTVEIPETNDVKLDYGNNPTEESEPQEGTPANQEIKVIKDWAVDGTITDANVAVKAIFTL
QEKQTDGTWVNVASHEATKPSRFEHTFTGLDNAKTYRVVERVSGYTPEYVSFKNGVVTIKNNKNSNDPTPINPSE
PKVVTYGRKFVKTNQANTERLAGATFLVKKEGKYLARKAGAATAEAKAAVKTAKLALDEAVKAYNDLTKEKQEGQ
EGKTALATVDQKQKAYNDAFVKANYSYEWVADKKADNVVKLISNAGGQFEITGLDKGTYGLEETQAPAGYATLSG
DVNFEVTATSYSKGATTDIAYDKGSVKKDAQQVQNKKVT*IPQTG*GIGTILFTIIGLSIMLGAVVIMKKRQSEEA

PC The GBS 59 polypeptides contain an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 129** IPQTG (shown in italics in SEQ ID NOs: 126 and 128 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 59 protein from the host cell. Alternatively, in some recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Pilin motifs, containing conserved lysine (K) residues have been identified in the GBS 59 polypeptides. The pilin motif sequences are underlined in each of SEQ ID NOs: 126 and 128, below. Conserved lysine (K) residues are marked in bold. The conserved lysine (K) residues are located at amino acid residues 202 and 212 and amino acid residues 489 and 495 of SEQ ID NO: 126 and at amino acid residues 188 and 198 of SEQ ID NO: 128. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of GBS 59. Preferred fragments of GBS 59 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 126

MKRINKYFAMFSALLLTLTSLLSVAPAFADEATTNTVTLHKILQTESNLNKSNFPGTTGLNGKDYKGGAISDLAG
YFGEGSKEIEGAFFALALKEDKSGKVQYVKAKEGNKLTPALINKDGTPEITVNIDEAVSGLTPEGDTGLVFNTKG
LKGEFKIVEVKSKSTYNNNGSLLAASKAVPVNITLPLVNEDGVVADAHVYPKNTEEKPEIDKNFAKTNDLTALTD
VNRLLTAGANYGNYARDKATATAEIGKVVPYEVKTKIHKGSKYENLVWTDIMSNGLTMGSTVSLKASGTTETFAK
DTDYELSIDARGFTLKFTADGLGKLEKAAKTADIEFTLTYSATVNGQAIIDNPESNDIKLSYGNKPGKDLTELPV
TPSKGEVTVAKTWSDGIAPDGVNVVYTLKDKDKTVASVSLTKTSKGTIDLGNGIKFEVSGNFSGKFTGLENKSYM
ISERVSGYGSAINLENGKVTITNTKDSDNPTPLNPTEPKVETHGKKFVKTNEQGDRLAGAQFVVKNSAGKYLALK
ADQSEGQKTLAAKKIALDEAIAAYNKLSATDQKGEKGITAKELIKTKQADYDAAFIEARTAYEWITDKARAITYT
SNDQGQFEVTGLADGTYNLEETLAPAGFAKLAGNIKFVVNQGSYITGGNIDYVANSNQKDATRVENKKVTIPQTG
GIGTILFTIIGLSIMLGAVVIMKRRQSKEA

SEQ ID NO: 128

MKKINKCLTMFSTLLLILTSLFSVAPAFADDATTDTVTLHKIVMPQAAFDNFTEGTKGKNDSDYVGKQINDLKSY
FGSTDAKEIKGAFFVFKNETGTKFITENGKEVDTLEAKDAEGGAVLSGLTKDNGFVFNTAKLKGIYQIVELKEKS
NYDNNGSILADSKAVPVKITLPLVNNQGVVKDAHIYPKNTETKPQVDKNFADKDLDYTDNRKDKGVVSATVGDKK
EYIVGTKILKGSDYKKLVWTDSMTKGLTFNNNVKVTLDGEDFPVLNYKLVTDDQGFRLALNATGLAAVAAAAKDK
DVEIKITYSATVNGSTTVEIPETNDVKLDYGNNPTEESEPQEGTPANQEIKVIKDWAVDGTITDANVAVKAIFTL
QEKQTDGTWVNVASHEATKPSRFEHTFTGLDNAKTYRVVERVSGYTPEYVSFKNGVVTIKNNKNSNDPTPINPSE
PKVVTYGRKFVKTNQANTERLAGATFLVKKEGKYLARKAGAATAEAKAAVKTAKLALDEAVKAYNDLTKEKQEGQ
EGKTALATVDQKQKAYNDAFVKANYSYEWVADKKADNVVKLISNAGGQFEITGLDKGTYGLEETQAPAGYATLSG
DVNFEVTATSYSKGATTDIAYDKGSVKKDAQQVQNKKVTIPQTGGIGTILFTIIGLSIMLGAVVIMKKRQSEEA

An E box containing a conserved glutamic residue has also been identified in each of the GBS 59 polypeptides. The E box motif is underlined in each of SEQ ID NOs: 126 and 128 below. The conserved glutamic acid (E) is marked in bold at amino acid residue 621 in SEQ ID NO: 126 and at amino acid residue 588 in SEQ ID NO: 128. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of GBS 59. Preferred fragments of GBS 59 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 126

MKRINKYFAMFSALLLTLTSLLSVAPAFADEATTNTVTLHKILQTESNLNKSNFPGTTGLNGKDYKGGAISDLAG
YFGEGSKEIEGAFFALALKEDKSGKVQYVKAKEGNKLTPALINKDGTPEITVNIDEAVSGLTPEGDTGLVFNTKG
LKGEFKIVEVKSKSTYNNNGSLLAASKAVPVNITLPLVNEDGVVADAHVYPKNTEEKPEIDKNFAKTNDLTALTD
VNRLLTAGANYGNYARDKATATAEIGKVVPYEVKTKIHKGSKYENLVWTDIMSNGLTMGSTVSLKASGTTETFAK
DTDYELSIDARGFTLKFTADGLGKLEKAAKTADIEFTLTYSATVNGQAIIDNPESNDIKLSYGNKPGKDLTELPV
TPSKGEVTVAKTWSDGIAPDGVNVVYTLKDKDKTVASVSLTKTSKGTIDLGNGIKFEVSGNFSGKFTGLENKSYM
ISERVSGYGSAINLENGKVTITNTKDSDNPTPLNPTEPKVETHGKKFVKTNEQGDRLAGAQFVVKNSAGKYLALK
ADQSEGQKTLAAKKIALDEAIAAYNKLSATDQKGEKGITAKELIKTKQADYDAAFIEARTAYEWITDKARAITYT
SNDQGQFEVTGLADGTYNLEETLAPAGFAKLAGNIKFVVNQGSYITGGNIDYVANSNQKDATRVENKKVTIPQTG
GIGTILFTIIGLSIMLGAVVIMKRRQSKEA

SEQ ID NO: 128

MKKINKCLTMFSTLLLILTSLFSVAPAFADDATTDTVTLHKIVMPQAAFDNFTEGTKGKNDSDYVGKQINDLKSY
FGSTDAKEIKGAFFVFKNETGTKFITENGKEVDTLEAKDAEGGAVLSGLTKDNGFVFNTAKLKGIYQIVELKEKS
NYDNNGSILADSKAVPVKITLPLVNNQGVVKDAHIYPKNTETKPQVDKNFADKDLDYTDNRKDKGVVSATVGDKK
EYIVGTKILKGSDYKKLVWTDSMTKGLTFNNNVKVTLDGEDFPVLNYKLVTDDQGFRLALNATGLAAVAAAAKDK
DVEIKITYSATVNGSTTVEIPETNDVKLDYGNNPTEESEPQEGTPANQEIKVIKDWAVDGTITDANVAVKAIFTL
QEKQTDGTWVNVASHEATKPSRFEHTFTGLDNAKTYRVVERVSGYTPEYVSFKNGVVTIKNNKNSNDPTPINPSE
PKVVTYGRKFVKTNQANTERLAGATFLVKKEGKYLARKAGAATAEAKAAVKTAKLALDEAVKAYNDLTKEKQEGQ
EGKTALATVDQKQKAYNDAFVKANYSYEWVADKKADNVVKLISNAGGQFEITGLDKGTYGLEETQAPAGYATLSG
DVNFEVTATSYSKGATTDIAYDKGSVKKDAQQVQNKKVTIPQTGGIGTILFTIIGLSIMLGAVVIMKKRQSEEA

Female mice were immunized with either SAG1407 (SEQ ID NO: 126) or BO1575 (SEQ ID NO: 128) in an active maternal immunization assay. Pups bred from the immunized female mice survived GBS challenge better than control (PBS) treated mice. Results of the active maternal immunization assay using the GBS 59 immunogenic compositions are shown in Table 17, below.

TABLE 17: Active maternal immunization assay for GBS 59

Challenge GBS strain (serotype)	GBS 59		PBS		
	Dead/treated	Survival (%)	Dead/treated	Survival (%)	FACS
CJB111 (V)*	7/20	65	41/49	16	493
18RS21 (II)**	18/30	40	39/40	2.5	380

* immunized with BO1575
**immunized with SAG1407

Opsonophagocytosis assays also demonstrated that antibodies against BO1575 are opsonic for GBS serotype V, strain CJB111. See Figure 67.

GBS 52

Examples of polynucleotide and amino acid sequences for GBS 52 are set forth below. SEQ ID NO: 20 and 21 represent GBS 52 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 20

ATGAAACAAACATTAAAACTTATGTTTTCTTTTCTGTTGATGTTAGGGACTATGTTTGGAATTAGCCAAACTGTT
TTAGCGCAAGAAACTCATCAGTTGACGATTGTTCATCTTGAAGCAAGGGATATTGATCGTCCAAATCCACAGTTG
GAGATTGCCCCTAAAGAAGGGACTCCAATTGAAGGAGTACTCTATCAGTTGTACCAATTAAAATCAACTGAAGAT
GGCGATTTGTTGGCACATTGGAATTCCCTAACTATCACAGAATTGAAAAAACAGGCGCAGCAGGTTTTTGAAGCC
ACTACTAATCAACAAGGAAAGGCTACATTTAACCAACTACCAGATGGAATTTATTATGGTCTGGCGGTTAAAGCC
GGTGAAAAAAATCGTAATGTCTCAGCTTTCTTGGTTGACTTGTCTGAGGATAAAGTGATTTATCCTAAAATCATC
TGGTCCACAGGTGAGTTGGACTTGCTTAAAGTTGGTGTGGATGGTGATACCAAAAAACCACTAGCAGGCGTTGTC
TTTGAACTTTATGAAAAGAATGGTAGGACTCCTATTCGTGTGAAAAATGGGGTGCATTCTCAAGATATTGACGCT
GCAAAACATTTAGAAACAGATTCATCAGGGCATATCAGAATTTCCGGGCTCATCCATGGGGACTATGTCTTAAAA
GAAATCGAGACACAGTCAGGATATCAGATCGGACAGGCAGAGACTGCTGTGACTATTGAAAAATCAAAAACAGTA

AGAGTAACGATTGAAAATAAAAAAGTTCCGACACCTAAAGTGCCATCTCGAGGAGGTCTTATTCCCAAAACAGGT
GAGCAACAGGCAATGGCACTTGTAATTATTGGTGGTATTTTAATTGCTTTAGCCTTACGATTACTATCAAAACAT
CGGAAACATCAAAATAAGGAT

SEQ ID NO: 21
MKQTLKLMFSFLLMLGTMFGISQTVLAQETHQLTIVHLEARDIDRPNPQLEIAPKEGTPIEGVLYQLYQLKSTED
GDLLAHWNSLTITELKKQAQQVFEATTNQQGKATFNQLPDGIYYGLAVKAGEKNRNVSAFLVDLSEDKVIYPKII
WSTGELDLLKVGVDGDTKKPLAGVVFELYEKNGRTPIRVKNGVHSQDIDAAKHLETDSSGHIRISGLIHGDYVLK
EIETQSGYQIGQAETAVTIEKSKTVTVTIENKKVPTPKVPSRGGL*IPKTG*EQQAMALVIIGGILIALALRLLSKH
RKHQNKD

GBS 52 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 124** IPKTG (shown in italics in SEQ ID NO: 21, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 52 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in GBS 52. The pilin motif sequence is underlined in SEQ ID NO: 21, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 148 and 160. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of GBS 52 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 21
MKQTLKLMFSFLLMLGTMFGISQTVLAQETHQLTIVHLEARDIDRPNPQLEIAPKEGTPIEGVLYQLYQLKSTED
GDLLAHWNSLTITELKKQAQQVFEATTNQQGKATFNQLPDGIYYGLAVKAGEKNRNVSAFLVDLSEDKVIYPKII
WSTGELDLLKVGVDGDTKKPLAGVVFELYEKNGRTPIRVKNGVHSQDIDAAKHLETDSSGHIRISGLIHGDYVLK
EIETQSGYQIGQAETAVTIEKSKTVTVTIENKKVPTPKVPSRGGLIPKTGEQQAMALVIIGGILIALALRLLSKH
RKHQNKD

An E box containing a conserved glutamic residue has been identified in GBS 52. The E-box motif is underlined in SEQ ID NO: 21, below. The conserved glutamic acid (E), at amino acid residue 226, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of GBS 52. Preferred fragments of GBS 52 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 21
MKQTLKLMFSFLLMLGTMFGISQTVLAQETHQLTIVHLEARDIDRPNPQLEIAPKEGTPIEGVLYQLYQLKSTED
GDLLAHWNSLTITELKKQAQQVFEATTNQQGKATFNQLPDGIYYGLAVKAGEKNRNVSAFLVDLSEDKVIYPKII
WSTGELDLLKVGVDGDTKKPLAGVVFELYEKNGRTPIRVKNGVHSQDIDAAKHLETDSSGHIRISGLIHGDYVLK
EIETQSGYQIGQAETAVTIEKSKTVTVTIENKKVPTPKVPSRGGLIPKTGEQQAMALVIIGGILIALALRLLSKH
RKHQNKD

SAG0647

Examples of polynucleotide and amino acid sequences for SAG0647 are set forth below.

SEQ ID NO: 22 and 23 represent SAG0647 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 22

ATGGGACAAAAATCAAAAATATCTCTAGCTACGAATATTCGTATATGGATTTTTCGTTTAATTTTCTTAGCGGGT
TTCCTTGTTTTGGCATTTCCCATCGTTAGTCAGGTCATGTACTTTCAAGCCTCTCACGCCAATATTAATGCTTTT
AAAGAAGCTGTTACCAAGATTGACCGGGTGGAGATTAATCGGCGTTTAGAACTTGCTTATGCTTATAACGCCAGT
ATAGCAGGTGCCAAAACTAATGGCGAATATCCAGCGCTTAAAGACCCCTACTCTGCTGAACAAAAGCAGGCAGGG
GTCGTTGAGTACGCCCGCATGCTTGAAGTCAAAGAACAAATAGGTCATGTGATTATTCCAAGAATTAATCAGGAT
ATCCCTATTTACGCTGGCTCTGCTGAAGAAAATCTTCAGAGGGGCGTTGGACATTTAGAGGGGACCAGTCTTCCA
GTCGGTGGTGAGTCAACTCATGCCGTTCTAACTGCCCATCGAGGGCTACCAACGGCCAAGCTATTTACCAATTTA
GACAAGGTAACAGTAGGTGACCGTTTTTACATTGAACACATCGGCGGAAAGATTGCTTATCAGGTAGACCAAATC
AAAGTTATCGCCCCTGATCAGTTAGAGGATTTGTACGTGATTCAAGGAGAAGATCACGTCACCCTATTAACTTGC
ACACCTTATATGATAAATAGTCATCGCCTCCTCGTTCGAGGCAAGCGAATTCCTTATGTGGAAAAAACAGTGCAG
AAAGATTCAAAGACCTTCAGGCAACAACAATACCTAACCTATGCTATGTGGGTAGTCGTTGGACTTATCTTGCTG
TCGCTTCTCATTTGGTTTAAAAAGACGAAACAGAAAAAGCGGAGAAAGAATGAAAAAGCGGCTAGTCAAAATAGT
CACAATAATTCGAAATAA

SEQ ID NO: 23

MGQKSKISLATNIRIWIFRLIFLAGFLVLAFPIVSQVMYFQASHANINAFKEAVTKIDRVEINRRLELAYAYNAS
IAGAKTNGEYPALKDPYSAEQKQAGVVEYARMLEVKEQIGHVIIPRINQDIPIYAGSAEENLQRGVGHLEGTSLP
VGGESTHAVLTAHRGLPTAKLFTNLDKVTVGDRFYIEHIGGKIAYQVDQIKVIAPDQLEDLYVIQGEDHVTLLTC
TPYMINSHRLLVRGKRIPYVEKTVQKDSKTFRQQQYLTYAMWVVVGLILLSLLIWFKKTKQKKRRKNEKAASQNS
HNNSK

SAG0648

Examples of polynucleotide and amino acid sequences for SAG0648 are set forth below.

SEQ ID NO: 24 and 25 represent SAG0648 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 24

ATGGGAAGTCTGATTCTCTTATTTCCGATTGTGAGCCAGGTAAGTTACTACCTTGCTTCGCATCAAAATATTAAT
CAATTTAAGCGGGAAGTCGCTAAGATTGATACTAATACGGTTGAACGACGCATCGCTTTAGCTAATGCTTACAAT
GAGACGTTATCAAGGAATCCCTTGCTTATAGACCCTTTTACCAGTAAGCAAAAAGAAGGTTTGAGAGAGTATGCT
CGTATGCTTGAAGTTCATGAGCAAATAGGTCATGTGGCAATCCCAAGTATTGGGGTTGATATTCCAATTTATGCT
GGAACATCCGAAACTGTGCTTCAGAAAGGTAGTGGGCATTTGGAGGGAACCAGTCTTCCAGTGGGAGGTTTGTCA
ACCCATTCAGTACTAACTGCCCACCGTGGCTTGCCAACAGCTAGGCTATTTACCGACTTAAATAAAGTTAAAAAA
GGCCAGATTTTCTATGTGACGAACATCAAGGAAACACTTGCCTACAAAGTCGTGTCTATCAAAGTTGTGGATCCA
ACAGCTTTAAGTGAGGTTAAGATTGTCAATGGTAAGGATTATATAACCTTGCTGACTTGCACACCTTACATGATC
AATAGTCATCGTCTCTTGGTAAAAGGAGAGCGTATTCCTTATGATTCTACCGAGGCGGAAAAGCACAAAGAACAA
ACCGTACAAGATTATCGTTTGTCACTAGTGTTGAAGATACTACTAGTATTATTAATTGGACTCTTCATCGTGATA
ATGATGAGAAGATGGATGCAACATCGTCAATAA

SEQ ID NO: 25

MGSLILLFPIVSQVSYYLASHQNINQFKREVAKIDTNTVERRIALANAYNETLSRNPLLIDPFTSKQKEGLREYA
RMLEVHEQIGHVAIPSIGVDIPIYAGTSETVLQKGSGHLEGTSLPVGGLSTHSVLTAHRGLPTARLFTDLNKVKK
GQIFYVTNIKETLAYKVVSIKVVDPTALSEVKIVNGKDYITLLTCTPYMINSHRLLVKGERIPYDSTEAEKHKEQ
TVQDYRLSLVLKILLVLLIGLFIVIMMRRWMQHRQ

GBS 150

Examples of polynucleotide and amino acid sequences for GBS 150 are set forth below. SEQ ID NO: 26 and 27 represent GBS 150 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 26

ATGAAAAAGATTAGAAAAAGTTTAGGACTTCTACTATGTTGCTTTTTAGGATTGGTACAATTAGCGTTTTTTTCG
GTAGCCAGTGTAAATGCTGATACCCCTAATCAACTAACAATCACACAGATAGGACTTCAGCCAAATACTACAGAG
GAGGGGATTTCTTATCGTTTATGGACTGTGACTGACAACTTAAAAGTTGATTTATTGAGCCAAATGACAGATAGC
GAATTGAACCAGAAGTATAAGAGTATCTTGACTTCTCCTACTGATACTAATGGTCAGACAAAGATAGCACTCCCA
AATGGTTCGTACTTTGGTCGTGCTTATAAAGCTGATCAAAGCGTTTCAACAATAGTACCTTTTTATATTGAATTA
CCAGATGATAAGTTATCAAATCAATTACAGATAAATCCTAAGCGAAAAGTTGAAACAGGCCGATTAAAACTTATT

AAATATACAAAAGAAGGAAAGATAAAGAAAAGGCTATCCGGAGTAATATTTGTATTATACGATAACCAGAATCAG
CCAGTTCGCTTTAAAAATGGACGATTTACGACCGATCAAGATGGGATTACTTCATTAGTAACTGATGATAAGGGA
GAAATTGAGGTTGAAGGTTTATTACCTGGTAAGTATATTTTTCGAGAAGCAAAAGCACTAACTGGTTACCGTATA
TCTATGAAGGATGCTGTAGTTGCTGTAGTTGCTAATAAAACACAGGAAGTAGAGGTAGAAAACGAAAAAGAAACT
CCTCCACCAACAAATCCTAAACCATCACAACCGCTTTTTCCACAATCATTTCTTCCTAAAACAGGAATGATTATT
GGTGGAGGACTGACAATTCTTGGTTGTATTATTTTGGGAATTTTGTTTATCTTTTTAAGAAAAACTAAAAATAGC
AAATCTGAAAGAAACGATACAGTA

SEQ ID NO: 27

MKKIRKSLGLLLCCFLGLVQLAFFSVASVNADTPNQLTITQIGLQPNTTEEGISYRLWTVTDNLKVDLLSQMTDS
ELNQKYKSILTSPTDTNGQTKIALPNGSYFGRAYKADQSVSTIVPFYIELPDDKLSNQLQINPKRKVETGRLKLI
KYTKEGKIKKRLSGVIFVLYDNQNQPVRFKNGRFTTDQDGITSLVTDDKGEIEVEGLLPGKYIFREAKALTGYRI
SMKDAVVAVVANKTQEVEVENEKETPPPTNPKPSQPLFPQSF*LPKTG*MIIGGGLTILGCIILGILFIFLRKTKNS
KSERNDTV

GBS 150 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 130** LPKTG (shown in italics in SEQ ID NO: 27 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GBS 150 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

As discussed above, a pilin motif, containing a conserved lysine (K) residue has been identified in GBS 150. The pilin motif sequence is underlined in SEQ ID NO: 27, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 139 and 148. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of GBS 150. Preferred fragments of GBS 150 include a conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 27

MKKIRKSLGLLLCCFLGLVQLAFFSVASVNADTPNQLTITQIGLQPNTTEEGISYRLWTVTDNLKVDLLSQMTDS
ELNQKYKSILTSPTDTNGQTKIALPNGSYFGRAYKADQSVSTIVPFYIELPDDKLSNQLQINPKRKVETGRLKLI
KYTKEGKIKKRLSGVIFVLYDNQNQPVRFKNGRFTTDQDGITSLVTDDKGEIEVEGLLPGKYIFREAKALTGYRI
SMKDAVVAVVANKTQEVEVENEKETPPPTNPKPSQPLFPQSFLPKTGMIIGGGLTILGCIILGILFIFLRKTKNS
KSERNDTV

An E box containing a conserved glutamic residue has also been identified in GBS 150. The E box motif is underlined in SEQ ID NO: 27 below. The conserved glutamic acid (E), at amino acid residue 216, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of GBS 150. Preferred fragments of GBS 150 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 27

MKKIRKSLGLLLCCFLGLVQLAFFSVASVNADTPNQLTITQIGLQPNTTEEGISYRLWTVTDNLKVDLLSQMTDS
ELNQKYKSILTSPTDTNGQTKIALPNGSYFGRAYKADQSVSTIVPFYIELPDDKLSNQLQINPKRKVETGRLKLI
KYTKEGKIKKRLSGVIFVLYDNQNQPVRFKNGRFTTDQDGITSLVTDDKGEIEVEGLLPGKYIFREAKALTGYRI
SMKDAVVAVVANKTQEVEVENEKETPPPTNPKPSQPLFPQSFLPKTGMIIGGGLTILGCIILGILFIFLRKTKNS
KSERNDTV

SAG1405

Examples of polynucleotide and amino acid sequences for SAG1405 are set forth below.

SEQ ID NO: 28 and 29 represent SAG1405 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 28

ATGGGAGGAAAATTTCAGAAAAACCTTAAGAAATCGGTCGTTTTAAATCGATGGATGAATGTAGGCTTGATACTA
TTGTTCTTAGTTGGTCTTTTGATAACCTCATATCCTTTTATTTCAAATTGGTACTATAATATTAAAGCTAATAAT
CAAGTAACTAACTTTGATAATCAAACCCAAAAATTAAATACTAAAGAGATTAATAGACGATTTGAGTTAGCAAAA
GCTTATAATAGAACACTGGACCCAAGCCGCCTATCAGATCCCTATACTGAAAAAGAAAAAAAAGGTATTGCTGAA
TACGCCCACATGCTTGAGATTGCTGAAATGATTGGATATATTGATATACCGTCTATCAAGCAAAAATTACCTATC
TATGCGGGGACTACCAGTAGTGTTCTTGAAAAAGGAGCAGGACACCTTGAAGGAACCTCCTTGCCAATTGGTGGA
AAAAGTTCACATACTGTTATCACAGCTCATCGCGGCTTACCTAAAGCTAAGTTATTTACAGATTTAGATAAACTT
AAAAAAGGAAAAATTTTTTATATTCATAATATCAAAGAAGTTTTAGCCTATAAGGTTGATCAAATAAGTGTTGTA
AAGCCAGATAATTTTTCTAAATTATTGGTTGTTAAAGGTAAGGATTATGCGACTTTGCTAACATGTACACCTTAT
TCGATTAATTCACATCGTTTACTAGTTAGAGGGCATCGAATCAAGTATGTACCTCCTGTTAAAGAAAAGAACTAT
TTAATGAAAGAATTGCAAACACACTATAAACTTTATTTCCTCTTATCAATCCTAGTTATTCTTATATTAGTCGCT
TTACTATTATATTTAAAACGAAAATTTAAAGAGAGAAAGAGAAAGGGAAATCAAAAATGA

SEQ ID NO: 29

MGGKFQKNLKKSVVLNRWMNVGLILLFLVGLLITSYPFISNWYYNIKANNQVTNFDNQTQKLNTKEINRRFELAK
AYNRTLDPSRLSDPYTEKEKKGIAEYAHMLEIAEMIGYIDIPSIKQKLPIYAGTTSSVLEKGAGHLEGTSLPIGG
KSSHTVITAHRGLPKAKLFTDLDKLKKGKIFYIHNIKEVLAYKVDQISVVKPDNFSKLLVVKGKDYATLLTCTPY
SINSHRLLVRGHRIKYVPPVKEKNYLMKELQTHYKLYFLLSILVILILVALLLYLKRKFKERKRKGNQK

SAG1406

Examples of polynucleotide and amino acid sequences for SAG1405 are set forth below.

SEQ ID NO: 30 and 31 represent SAG1405 sequences from GBS serotype V, strain isolate 2603.

SEQ ID NO: 30

GTGAAGACTAAAAAAATCATCAAAAAAACAAAAAAAAAGAAGAAGTCAAATCTTCCTTTTATCATTCTTTTTCTA
ATAGGTCTATCTATTTTATTGTATCCAGTGGTATCACGTTTTTACTATACGATAGAATCTAATAATCAAACACAG
GATTTTGAGAGAGCTGCTAAAAAACTTAGTCAGAAAGAAATCAATCGACGTATGGCTCTAGCACAAGCTTATAAT
GATTCTTTAAATAATGTCCATCTTGAAGATCCTTATGAGAAAAAACGAATTCAAAAGGGGGTAGCAGAGTACGCC
CGTATGTTAGAGGTAAGTGAAAAAATCGGAACAATTTCAGTTCCTAAGATAGGTCAAAAACTCCCTATATTTGCA
GGTTCAAGTCAAGAAGTTCTATCTAAAGGAGCAGGGCATTTAGAAGGTACCTCTCTTCCAATTGGGGGCAATAGT
ACACATACTGTTATAACAGCGCATTCAGGAATTCCAGATAAAGAACTCTTTTCTAACCTTAAAAAGTTAAAAAAA
GGAGATAAGTTTTATATTCAAAACATAAAAGAAACGATAGCATATCAAGTAGATCAGATAAAAGTCGTTACACCC
GATAACTTTTCAGATTTGTTGGTTGTTCCTGGACATGATTATGCAACCTTATTGACTTGCACCCCGATTATGATC
AATACACACAGACTTTTAGTAAGGGGACATCGTATCCCTTATAAAGGTCCTATTGATGAAAAATTAATAAAAGAC
GGTCATTTAAACACGATTTATAGATATCTATTCTATATATCTTTAGTTATTATTGCTTGGTTACTTTGGTTAATA
AAACGTCAACGTCAAAAAAATCGTTTAGCAAGTGTTAGAAAAGGAATTGAATCATAA

SEQ ID NO: 31

MKTKKIIKKTKKKKKSNLPFIILFLIGLSILLYPVVSRFYYTIESNNQTQDFERAAKKLSQKEINRRMALAQAYN
DSLNNVHLEDPYEKKRIQKGVAEYARMLEVSEKIGTISVPKIGQKLPIFAGSSQEVLSKGAGHLEGTSLPIGGNS
THTVITAHSGIPDKELFSNLKKLKKGDKFYIQNIKETIAYQVDQIKVVTPDNFSDLLVVPGHDYATLLTCTPIMI
NTHRLLVRGHRIPYKGPIDEKLIKDGHLNTIYRYLFYISLVIIAWLLWLIKRQRQKNRLASVRKGIES

01520

An example of an amino acid sequence for 01520 is set forth below. SEQ ID NO: 32 represents a 01520 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 32

MIRRYSANFLAILGIILVSSGIYWGWYNINQAHQADLTSQHIVKVLDKSITHQVKGSENGELPVKKLDKTDYLGT
LDIPNLKLHLPVAANYSFEQLSKTPTRYYGSYLTNNMVICAHNFPYHFDALKNVDMGTDVYFTTTTGQIYHYKIS
NREIIEPTAIEKVYKTATSDNDWDLSLFTCTKAGVARVLVRCQLIDVKN

01521

An example of an amino acid sequence for 01521 is set forth below. SEQ ID NO: 33 represents a 01521 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 33

MIYKKILKITLLLLFSLSTQLVSADTNDQMKTGSITIQNKYNNQGIAGGNLLVYQVAQAKDVDGNQVFTLTTPFQ
GIGIKDDDLTQVNLDSNQAKYVNLLTKAVHKTQPLQTFDNLPAEGIVANNLPQGIYLFIQTKTAQGYELMSPFIL
SIPKDGKYDITAFEKMSPLNAKPKKEETITPTVTHQTKGK*LPFTG*QVWWPIPILIMSGLLCLIIALKWRRRRD

01521 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 132** LPFTG (shown in italics in SEQ ID NO: 33 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 01521 protein from the host cell. Alternatively, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, containing conserved lysine (K) residues have been identified in 01521. The pilin motif sequences are underlined in SEQ ID NO: 33, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 154 and 165 and at amino acid residues 174 and 188. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of 01521. Preferred fragments of 01521 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 33

MIYKKILKITLLLLFSLSTQLVSADTNDQMKTGSITIQNKYNNQGIAGGNLLVYQVAQAKDVDGNQVFTLTTPFQ
GIGIKDDDLTQVNLDSNQAKYVNLLTKAVHKTQPLQTFDNLPAEGIVANNLPQGIYLFIQTKTAQGYELMSPFIL
SIPKDGKYDITAFEKMSPLNAKPKKEETITPTVTHQTKGKLPFTGQVWWPIPILIMSGLLCLIIALKWRRRRD

An E box containing a conserved glutamic residue has also been identified in 01521. The E box motif is underlined in SEQ ID NO: 33 below. The conserved glutamic acid (E), at amino acid residue 177, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of 01521. Preferred fragments of 01521 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 33

MIYKKILKITLLLLFSLSTQLVSADTNDQMKTGSITIQNKYNNQGIAGGNLLVYQVAQAKDVDGNQVFTLTTPFQ
GIGIKDDDLTQVNLDSNQAKYVNLLTKAVHKTQPLQTFDNLPAEGIVANNLPQGIYLFIQTKTAQGYELMSPFIL
SIPKDGKYDITAFEKMSPLNAKPKKEETITPTVTHQTKGKLPFTGQVWWPIPILIMSGLLCLIIALKWRRRRD

01522

An example of an amino acid sequence for 01522 is set forth below. SEQ ID NO: 34 represents a 01522 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 34

MAYPSLANYWNSFHQSRAIMDYQDRVTHMDENDYKKIINRAKEYNKQFKTSGMKWHMTSQERLDYNSQLAIDKTG
NMGYISIPKINIKLPLYHGTSEKVLQTSIGHLEGSSLPIGGDSTHSILSGHRGLPSSRLFSDLDKLKVGDHWTVS
ILNETYTYQVDQIRTVKPDDLRDLQIVKGKDYQTLVTCTPYGVNTHRLLVRGHRVPNDNGNALVVAEAIQIEPIY
IAPFIAIFLTLILLLLISLEVTRRARQRKKILKQAMRKEENNDL

01523

An example of an amino acid sequence for 01523 is set forth below. SEQ ID NO: 35 represents a 01523 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 35

MKKKMIQSLLVASLAFGMAVSPVTPIAFAAETGTITVQDTQKGATYKAYKVFDAEIDNANVSDSNKDGASYLIPQ
GKEAEYKASTDFNSLFTTTTNGGRTYVTKKDTASANEIATWAKSISANTTPVSTVTESNNDGTEVINVSQYGYYY
VSSTVNNGAVIMVTSVTPNATIHEKNTDATWGDGGGKTVDQKTYSVGDTVKYTITYKNAVNYHGTEKVYQYVIKD
TMPSASVVDLNEGSYEVTITDGSGNITTLTQGSEKATGKYNLLEENNNFTITIPWAATNTPTGNTQNGANDDFFY
KGINTITVTYTGVLKSGAKPGSADLPENTNIATINPNTSNDDPGQKVTVRDGQITIKKIDGSTKASLQGAIFVLK
NATGQFLNFNDTNNVEWGTEANATEYTTGADGIITITGLKEGTYYLVEKKAPLGYNLLDNSQKVILGDGATDTTN
SDNLLVNPTVENNKGTE*LPSTG*GIGTTIFYIIGAILVIGAGIVLVARRRLRS

01523 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 131** LPSTG (shown in italics in SEQ ID NO: 35 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 01523 protein from the host cell. Alternatively, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

An E box containing a conserved glutamic residue has also been identified in 01523. The E box motif is underlined in SEQ ID NO: 35 below. The conserved glutamic acid (E), at amino acid residue 423, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of 01523. Preferred fragments of 01523 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 35

MKKKMIQSLLVASLAFGMAVSPVTPIAFAAETGTITVQDTQKGATYKAYKVFDAEIDNANVSDSNKDGASYLIPQ
GKEAEYKASTDFNSLFTTTTNGGRTYVTKKDTASANEIATWAKSISANTTPVSTVTESNNDGTEVINVSQYGYYY
VSSTVNNGAVIMVTSVTPNATIHEKNTDATWGDGGGKTVDQKTYSVGDTVKYTITYKNAVNYHGTEKVYQYVIKD
TMPSASVVDLNEGSYEVTITDGSGNITTLTQGSEKATGKYNLLEENNNFTITIPWAATNTPTGNTQNGANDDFFY
KGINTITVTYTGVLKSGAKPGSADLPENTNIATINPNTSNDDPGQKVTVRDGQITIKKIDGSTKASLQGAIFVLK
NATGQFLNFNDTNNVEWGTEANATEYTTGADGIITITGLKEGT<u>YYLVEKKAPLGY</u>NLLDNSQKVILGDGATDTTN
SDNLLVNPTVENNKGTELPSTGGIGTTIFYIIGAILVIGAGIVLVARRRLRS

<u>01524</u>

An example of an amino acid sequence for 01524 is set forth below. SEQ ID NO: 36 represents a 01524 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 36

MLKKCQTFIIESLKKKKHPKEWKIIMWSLMILTTFLTTYFLILPAITVEETKTDDVGITLENKNSSQVTSSTSSS
QSSVEQSKPQTPASSVTETSSSEEAAYREEPLMFRGADYTVTVTLTKEAKIPKNADLKVTELKDNSATFKDYKKK
ALTEVAKQDSEIKNFKLYDITIESNGKEAEPQAPVKVEVNYDKPLEASDENLKVVHFKDDGQTEVLKSKDTAETK
NTSSDVAFKTDSFSIYAIVQEDNTEVPRLTYHFQNNDGTDYDFLTASGMQVHHQIIKDGESLGEVGIPTIKAGEH
FNGWYTYDPTTGKYGDPVKFGEPITVTETKEICVRPFMSKVATVTLYDDSAGKSILERYQVPLDSSGNGTADLSS
FKVSPPTSTLLFVGWSKTQNGAPLSESEIQALPVSSDISLYPVFKESYGVEFNTGDLSTGVTYIAPRRVLTGQPA
STIKPNDPTRPGYTFAGWYTAASGGAAFDFNQVLTKDTTLYAHWSPAQTTYTINYWQQSATDNKNATDAQKTYEY
AGQVTRSGLSLSNQTLTQQDINDKLPTGFKVNNTRTETSVMIKDDGSSVVNVYYDRKLITIKFAKYGGYSLPEYY
YSYNWSSDADTYTGLYGTTLAANGYQWKTGAWGYLANVGNNQVGTYGMSYLGEFILPNDTVDSDVIKLFPKGNIV
QTYRFFKQGLDGTYSLADTGGGAGADEFTFTEKYLGFNVKYYQRLYPDNYLFDQYASQTSAGVKVPISDEYYDRY
GAYHKDYLNLVVWYERNSYKIKYLDPLDNTELPNFPVKDVLYEQNLSSYAPDTTTVQPKPSRPGYVWDGKWYKDQ
AQTQVFDFNTTMPPHDVKVYAGWQKVTYRVNIDPNGGRLSKTDDTYLDLHYGDRIPDYTDITRDYIQDPSGTYYY
KYDSRDKDPDSTKDAYYTTDTSLSNVDTTTKYKYVKDAYKLVGWYYVNPDGSIRPYNFSGAVTQDINLRAIWRKA

GDYHIIYSNDAVGTDGKPALDASGQQLQTSNEPTDPDSYDDGSHSALLRRPTMPDGYRFRGWWYNGKIYNPYDSI
DIDAHLADANKNITIKPVIIPVGDIKLEDTSIKYNGNGGTRVENGNVVTQVETPRMELNSTTTIPENQYFTRTGY
NLIGWHHDKDLADTGRVEFTAGQSIGIDNNPDATNTLYAVWQPKEYTVRVSKTVVGLDEDKTKDFLFNPSETLQQ
ENFPLRDGQTKEFKVPYGTSISIDEQAYDEFKVSESITEKNLATGEADKTYDATGLQSLTVSGDVDISFTNTRIK
QKVRLQKVNVENDNNFLAGAVFDIYESDANGNKASHPMYSGLVTNDKGLLLVDANNYLSLPVGKYYLTETKAPPG
YLLPKNDISVLVISTGVTFEQNGNNATPIKENLVDGSTVYTFKITNSKGTE*LPSTG*GIGTHIYILVGLALALPSG
LILYYRKKI

01524 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 131** LPSTG (shown in italics in SEQ ID NO: 36 above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 01524 protein from the host cell. Alternatively, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Three pilin motifs, containing conserved lysine (K) residues have been identified in 01524. The pilin motif sequences are underlined in SEQ ID NO: 36, below. Conserved lysine (K) residues are marked in bold, at amino acid residues 128 and 138, amino acid residues 671 and 682, and amino acid residues 809 and 820. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures of 01524. Preferred fragments of 01524 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 36

MLKKCQTFIIESLKKKKHPKEWKIIMWSLMILTTFLTTYFLILPAITVEETKTDDVGITLENKNSSQVTSSTSSS
QSSVEQSKPQTPASSVTETSSSEEAAYREEPLMFRGADYTVTVTLTKEAKIPKNADLKVTELKDNSATFKDYKKK
ALTEVAKQDSEIKNFKLYDITIESNGKEAEPQAPVKVEVNYDKPLEASDENLKVVHFKDDGQTEVLKSKDTAETK
NTSSDVAFKTDSFSIYAIVQEDNTEVPRLTYHFQNNDGTDYDFLTASGMQVHHQIIKDGESLGEVGIPTIKAGEH
FNGWYTYDPTTGKYGDPVKFGEPITVTETKEICVRPFMSKVATVTLYDDSAGKSILERYQVPLDSSGNGTADLSS
FKVSPPTSTLLFVGWSKTQNGAPLSESEIQALPVSSDISLYPVFKESYGVEFNTGDLSTGVTYIAPRRVLTGQPA
STIKPNDPTRPGYTFAGWYTAASGGAAFDFNQVLTKDTTLYAHWSPAQTTYTINYWQQSATDNKNATDAQKTYEY
AGQVTRSGLSLSNQTLTQQDINDKLPTGFKVNNTRTETSVMIKDDGSSVVNVYYDRKLITIKFAKYGGYSLPEYY
YSYNWSSDADTYTGLYGTTLAANGYQWKTGAWGYLANVGNNQVGTYGMSYLGEFILPNDTVDSDVIKLFPKGNIV
QTYRFFKQGLDGTYSLADTGGGAGADEFTFTEKYLGFNVKYYQRLYPDNYLFDQYASQTSAGVKVPISDEYYDRY
GAYHKDYLNLVVWYERNSYKIKYLDPLDNTELPNFPVKDVLYEQNLSSYAPDTTTVQPKPSRPGYVWDGKWYKDQ
AQTQVFDFNTTMPPHDVKVYAGWQKVTYRVNIDPNGGRLSKTDDTYLDLHYGDRIPDYTDITRDYIQDPSGTYYY
KYDSRDKDPDSTKDAYYTTDTSLSNVDTTTKYKYVKDAYKLVGWYYVNPDGSIRPYNFSGAVTQDINLRAIWRKA
GDYHIIYSNDAVGTDGKPALDASGQQLQTSNEPTDPDSYDDGSHSALLRRPTMPDGYRFRGWWYNGKIYNPYDSI
DIDAHLADANKNITIKPVIIPVGDIKLEDTSIKYNGNGGTRVENGNVVTQVETPRMELNSTTTIPENQYFTRTGY
NLIGWHHDKDLADTGRVEFTAGQSIGIDNNPDATNTLYAVWQPKEYTVRVSKTVVGLDEDKTKDFLFNPSETLQQ
ENFPLRDGQTKEFKVPYGTSISIDEQAYDEFKVSESITEKNLATGEADKTYDATGLQSLTVSGDVDISFTNTRIK
QKVRLQKVNVENDNNFLAGAVFDIYESDANGNKASHPMYSGLVTNDKGLLLVDANNYLSLPVGKYYLTETKAPPG
YLLPKNDISVLVISTGVTFEQNGNNATPIKENLVDGSTVYTFKITNSKGTELPSTGGIGTHIYILVGLALALPSG
LILYYRKKI

An E box containing a conserved glutamic residue has also been identified in 01524. The E box motif is underlined in SEQ ID NO: 36 below. The conserved glutamic acid (E), at amino acid residue 1344, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of 01524. Preferred

fragments of 01324 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 36

MLKKCQTFIIESLKKKKHPKEWKIIMWSLMILTTFLTTYFLILPAITVEETKTDDVGITLENKNSSQVTSSTSSS
QSSVEQSKPQTPASSVTETSSSEEAAYREEPLMFRGADYTVTVTLTKEAKIPKNADLKVTELKDNSATFKDYKKK
ALTEVAKQDSEIKNFKLYDITIESNGKEAEPQAPVKVEVNYDKPLEASDENLKVVHFKDDGQTEVLKSKDTAETK
NTSSDVAFKTDSFSIYAIVQEDNTEVPRLTYHFQNNDGTDYDFLTASGMQVHHQIIKDGESLGEVGIPTIKAGEH
FNGWYTYDPTTGKYGDPVKFGEPITVTETKEICVRPFMSKVATVTLYDDSAGKSILERYQVPLDSSGNGTADLSS
FKVSPPTSTLLFVGWSKTQNGAPLSESEIQALPVSSDISLYPVFKESYGVEFNTGDLSTGVTYIAPRRVLTGQPA
STIKPNDPTRPGYTFAGWYTAASGGAAFDFNQVLTKDTTLYAHWSPAQTTYTINYWQQSATDNKNATDAQKTYEY
AGQVTRSGLSLSNQTLTQQDINDKLPTGFKVNNTRTETSVMIKDDGSSVVNVYYDRKLITIKFAKYGGYSLPEYY
YSYNWSSDADTYTGLYGTTLAANGYQWKTGAWGYLANVGNNQVGTYGMSYLGEFILPNDTVDSDVIKLFPKGNIV
QTYRFFKQGLDGTYSLADTGGGAGADEFTFTEKYLGFNVKYYQRLYPDNYLFDQYASQTSAGVKVPISDEYYDRY
GAYHKDYLNLVVWYERNSYKIKYLDPLDNTELPNFPVKDVLYEQNLSSYAPDTTTVQPKPSRPGYVWDGKWYKDQ
AQTQVFDFNTTMPPHDVKVYAGWQKVTYRVNIDPNGGRLSKTDDTYLDLHYGDRIPDYTDITRDYIQDPSGTYYY
KYDSRDKDPDSTKDAYYTTDTSLSNVDTTTKYKYVKDAYKLVGWYYVNPDGSIRPYNFSGAVTQDINLRAIWRKA
GDYHIIYSNDAVGTDGKPALDASGQQLQTSNEPTDPDSYDDGSHSALLRRPTMPDGYRFRGWWYNGKIYNPYDSI
DIDAHLADANKNITIKPVIIPVGDIKLEDTSIKYNGNGGTRVENGNVVTQVETPRMELNSTTTIPENQYFTRTGY
NLIGWHHDKDLADTGRVEFTAGQSIGIDNNPDATNTLYAVWQPKEYTVRVSKTVVGLDEDKTKDFLFNPSETLQQ
ENFPLRDGQTKEFKVPYGTSISIDEQAYDEFKVSESITEKNLATGEADKTYDATGLQSLTVSGDVDISFTNTRIK
QKVRLQKVNVENDNNFLAGAVFDIYESDANGNKASHPMYSGLVTNDKGLLLVDANNYLSLPVGKYYLTETKAPPG
YLLPKNDISVLVISTGVTFEQNGNNATPIKENLVDGSTVYTFKITNSKGTELPSTGGIGTHIYILVGLALALPSG
LILYYRKKI

01525

An example of an amino acid sequence for 01525 is set forth below. SEQ ID NO: 37 represents a 01525 sequence from GBS serotype III, strain isolate COH1.

SEQ ID NO: 37

MKRQISSDKLSQELDRVTYQKRFWSVIKNTIYILMAVASIAILIAVLWLPVLRIYGHSMNKTLSAGDVVFTVKGS
NFKTGDVVAFYYNNKVLVKRVIAESGDWVNIDSQGDVYVNQHKLKEPYVIHKALGNSNIKYPYQVPDKKIFVLGD
NRKTSIDSRSTSVGDVSEEQIVGKISFRIWPLGKISSIN

GBS 322

GBS 322 refers to a surface immunogenic protein, also referred to as "sip". Nucleotide and amino acid sequences of GBS 322 sequenced from serotype V isolated strain 2603 V/R are set forth in Ref. 3 as SEQ ID 8539 and SEQ ID 8540. These sequences are set forth below as SEQ ID NOS 38 and 39:

SEQ ID NO. 38

ATGAATAAAAAGGTACTATTGACATCGACAATGGCAGCTTCGCTATTATCAGTCGCAAGTGTTCAAGCACAAGAA
ACAGATACGACGTGGACAGCACGTACTGTTTCAGAGGTAAAGGCTGATTTGGTAAAGCAAGACAATAAATCATCA
TATACTGTGAAATATGGTGATACACTAAGCGTTATTTCAGAAGCAATGTCAATTGATATGAATGTCTTAGCAAAA
ATAAATAACATTGCAGATATCAATCTTATTTATCCTGAGACAACACTGACAGTAACTTACGATCAGAAGAGTCAT
ACTGCCACTTCAATGAAAATAGAAACACCAGCAACAAATGCTGCTGGTCAAACAACAGCTACTGTGGATTTGAAA
ACCAATCAAGTTTCTGTTGCAGACCAAAAAGTTTCTCTCAATACAATTTCGGAAGGTATGACACCAGAAGCAGCA
ACAACGATTGTTTCGCCAATGAAGACATATTCTTCTGCGCCAGCTTTGAAATCAAAAGAAGTATTAGCACAAGAG
CAAGCTGTTAGTCAAGCAGCAGCTAATGAACAGGTATCACCAGCTCCTGTGAAGTCGATTACTTCAGAAGTTCCA
GCAGCTAAAGAGGAAGTTAAACCAACTCAGACGTCAGTCAGTCAGTCAACAACAGTATCACCAGCTTCTGTTGCC
GCTGAAACACCAGCTCCAGTAGCTAAAGTAGCACCGGTAAGAACTGTAGCAGCCCCTAGAGTGGCAAGTGTTAAA
GTAGTCACTCCTAAAGTAGAAACTGGTGCATCACCAGAGCATGTATCAGCTCCAGCAGTTCCTGTGACTACGACT
TCACCAGCTACAGACAGTAAGTTACAAGCGACTGAAGTTAAGAGCGTTCCGGTAGCACAAAAAGCTCCAACAGCA
ACACCGGTAGCACAACCAGCTTCAACAACAAATGCAGTAGCTGCACATCCTGAAAATGCAGGGCTCCAACCTCAT
GTTGCAGCTTATAAAGAAAAAGTAGCGTCAACTTATGGAGTTAATGAATTCAGTACATACCGTGCGGGAGATCCA
GGTGATCATGGTAAAGGTTTAGCAGTTGACTTTATTGTAGGTACTAATCAAGCACTTGGTAATAAAGTTGCACAG
TACTCTACACAAAATATGGCAGCAAATAACATTTCATATGTTATCTGGCAACAAAAGTTTTACTCAAATACAAAC
AGTATTTATGGACCTGCTAATACTTGGAATGCAATGCCAGATCGTGGTGGCGTTACTGCCAACCACTATGACCAC

GTTCAGGTATCAATTTAACAAATAAATATAAAAAAAGGAAGCTATTTGGCTTCTTTTTTATATGCCTTGAATAGACTT
TCAAGGTTCTTATATAATTTTTATTA

SEQ ID NO. 39

MNKKVLLTSTMAASLLSVASVQAQETDTTWTARTVSEVKADLVKQDNKSSYTVKYGDTLSVISEAMSIDMNVLAK
INNIADINLIYPETTLTVTYDQKSHTATSMKIETPATNAAGQTTATVDLKTNQVSVADQKVSLNTISEGMTPEAA
TTIVSPMKTYSSAPALKSKEVLAQEQAVSQAAANEQVSPAPVKSITSEVPAAKEEVKPTQTSVSQSTTVSPASVA
AETPAPVAKVAPVRTVAAPRVASVKVVTPKVETGASPEHVSAPAVPVTTTSPATDSKLQATEVKSVPVAQKAPTA
TPVAQPASTTNAVAAHPENAGLQPHVAAYKEKVASTYGVNEFSTYRAGDPGDHGKGLAVDFIVGTNQALGNKVAQ
YSTQNMAANNISYVIWQQKFYSNTNSIYGPANTWNAMPDRGGVTANHYDHVHVSFNK

GBS 322 contains an N-terminal leader or signal sequence region which is indicated by the underlined sequence near the beginning of SEQ ID NO: 39. In one embodiment, one or more amino acids from the leader or signal sequence region of GBS 322 are removed. An example of such a GBS 322 fragment is set forth below as SEQ ID NO: 40.

SEQ ID NO: 40

DLVKQDNKSSYTVKYGDTLSVISEAMSIDMNVLAKINNIADINLIYPETTLTVTYDQKSHTATSMKIETPATNAA
GQTTATVDLKTNQVSVADQKVSLNTISEGMTPEAATTIVSPMKTYSSAPALKSKEVLAQEQAVSQAAANEQVSPA
PVKSITSEVPAAKEEVKPTQTSVSQSTTVSPASVAAETPAPVAKVAPVRTVAAPRVASVKVVTPKVETGASPEHV
SAPAVPVTTTSPATDSKLQATEVKSVPVAQKAPTATPVAQPASTTNAVAAHPENAGLQPHVAAYKEKVASTYGVN
EFSTYRAGDPGDHGKGLAVDFIVGTNQALGNKVAQYSTQNMAANNISYVIWQQKFYSNTNSIYGPANTWNAMPDR
GGVTANHYDHVHVSFNK

Additional preferred fragments of GBS 322 comprise the immunogenic epitopes identified in WO 03/068813, each of which are specifically incorporated by reference herein.

There may be an upper limit to the number of GBS proteins which will be in the compositions of the invention. Preferably, the number of GBS proteins in a composition of the invention is less than 20, less than 19, less than 18, less than 17, less than 16, less than 15, less than 14, less than 13, less than 12, less than 11, less than 10, less than 9, less than 8, less than 7, less than 6, less than 5, less than 4, or less than 3. Still more preferably, the number of GBS proteins in a composition of the invention is less than 6, less than 5, or less than 4. Still more preferably, the number of GBS proteins in a composition of the invention is 3.

The GBS proteins and polynucleotides used in the invention are preferably isolated, *i.e.*, separate and discrete, from the whole organism with which the molecule is found in nature or, when the polynucleotide or polypeptide is not found in nature, is sufficiently free of other biological macromolecules so that the polynucleotide or polypeptide can be used for its intended purpose.

Group A Streptococcus Adhesin Island Sequences

The GAS AI polypeptides of the invention can, of course, be prepared by various means (*e.g.* recombinant expression, purification from GAS, chemical synthesis *etc.*) and in various forms (*e.g.* native, fusions, glycosylated, non-glycosylated *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other streptococcal or host cell proteins) or substantially isolated form.

The GAS AI proteins of the invention may include polypeptide sequences having sequence identity to the identified GAS proteins. The degree of sequence identity may vary depending on the amino acid sequence (a) in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%,

80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more). Polypeptides having sequence identity include homologs, orthologs, allelic variants and functional mutants of the identified GBS proteins. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affinity gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The GAS adhesin island polynucleotide sequences may include polynucleotide sequences having sequence identity to the identified GAS adhesin island polynucleotide sequences. The degree of sequence identity may vary depending on the polynucleotide sequence in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more).

The GAS adhesin island polynucleotide sequences of the invention may include polynucleotide fragments of the identified adhesin island sequences. The length of the fragment may vary depending on the polynucleotide sequence of the specific adhesin island sequence, but the fragment is preferably at least 10 consecutive polynucleotides, (*e.g.* at least 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more).

The GAS adhesin island amino acid sequences of the invention may include polypeptide fragments of the identified GAS proteins. The length of the fragment may vary depending on the amino acid sequence of the specific GAS antigen, but the fragment is preferably at least 7 consecutive amino acids, (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). Preferably the fragment comprises one or more epitopes from the sequence. Other preferred fragments include (1) the N-terminal signal peptides of each identified GAS protein, (2) the identified GAS protein without their N-terminal signal peptides, and (3) each identified GAS protein wherein up to 10 amino acid residues (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) are deleted from the N-terminus and/or the C-terminus *e.g.* the N-terminal amino acid residue may be deleted. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

GAS AI-1 sequences

As discussed above, a GAS AI-1 sequence is present in an M6 strain isolate (MGAS10394). Examples of GAS AI-1 sequences from M6 strain isolate MGAS10394 are set forth below.

M6_Spy0156: Spy0156 is a rofA transcriptional regulator. An example of an amino acid sequence for M6_Spy0156 is set forth in SEQ ID NO: 41.

SEQ ID NO: 41

MIEKYLESSIESKCQLVVLFFKTSYLPITEVAEKTGLTFLQLNHYCEELNAFFPDSLSMTIQKRMISCQFTHPFK
ETYLYQLYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNSSAYRMREALIPLLRNFELKLSKNKIVGEEYRIR
YLIALLYSKFGIKVYDLTQQDKNTIHSFLSHSSTHLKTSPWLSESFSFYDILLALSWKRHQFSVTIPQTRIFQQL
KKLFIYDSLKKSSRDIIETYCQLNFSAGDLDYLYLIYITANNSFASLQWTPEHIRQCCQLFEENDTFRLLLKPII
TLLPNLKEQKPSLVKALMFFSKSFLFNLQHFIPETNLFVSPYYKGNQKLYTSLKLIVEEWLAKLPGKRYLNHKHF
HLFCHYVEQILRNIQPPLVVVFVASNFINAHLLTDSFPRYFSDKSIDFHSYIAR

M6_Spy0157: M6_Spy0157 is a fibronectin binding protein. It contains a sortase substrate motif LPXTG (SEQ ID NO: 122), shown in italics in the amino acid sequence SEQ ID NO: 42 .

SEQ ID NO: 42

MVSSYMFVRGEKMNNKIFLNKEASFLAHTKRKRRFAVTLVGVFFMLLACAGAIGFGQVAYAADEKTVPSHSSPNP
EFPWYGYDAYGKEYPGYNIWTRYHDLRVNLNGSRSYQVYCFNIQSNYPSQKNSFIKNWFKKIEGNGKSFVDYAHT
TKLGKEELEQRLLSLLYNAYPNDANGYMKGLEHLNAITVTQYAVWHYSDNSQYQFETLWESEAKEGKISRSQVTL
MREALKKLIDPNLEATAVNKIPSGYRLNIFESENEAYQNLLSAEYVPDDPPKPGETSEHNPKTPELDGTPIPEDP
KHPDDNLEPTLPPVMLDGEEVPEVPSESLEPALPPLMPELDGQEVPEKPSIDLPIEVPRYEFNNKDQSPLAGESG
ETEYITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDT
KEPEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTK
EPEVLMGGQSESVEFTKDTQTGMSGFSETATVVEDTRPKLVFHFDNNEPKVEENREKPTKNITPI*LPATG*DIENV
LAFLGILILSVLSIFSLLKNKQSNKKV

M6_Spy0157 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 180** LPATG (shown in italics in SEQ ID NO: 42, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant M6_Spy0157 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in M6_Spy0157. The pilin motif sequence is underlined in SEQ ID NO: 42, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 277, 287, and 301. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of M6_Spy0157 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 42

MVSSYMFVRGEKMNNKIFLNKEASFLAHTKRKRRFAVTLVGVFFMLLACAGAIGFGQVAYAADEKTVPSHSSPNP
EFPWYGYDAYGKEYPGYNIWTRYHDLRVNLNGSRSYQVYCFNIQSNYPSQKNSFIKNWFKKIEGNGKSFVDYAHT
TKLGKEELEQRLLSLLYNAYPNDANGYMKGLEHLNAITVTQYAVWHYSDNSQYQFETLWESEAKEGKISRSQVTL
MREALKKLIDPNLEATAVNKIPSGYRLNIFESENEAYQNLLSAEYVPDDPPKPGETSEHNPKTPELDGTPIPEDP
KHPDDNLEPTLPPVMLDGEEVPEVPSESLEPALPPLMPELDGQEVPEKPSIDLPIEVPRYEFNNKDQSPLAGESG
ETEYITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDT
KEPEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTK
EPEVLMGGQSESVEFTKDTQTGMSGFSETATVVEDTRPKLVFHFDNNEPKVEENREKPTKNITPILPATGDIENV
LAFLGILILSVLSIFSLLKNKQSNKKV

A repeated series of four E boxes containing a conserved glutamic residue have been identified in M6_Spy0157. The E-box motifs are underlined in SEQ ID NO: 42, below. The conserved glutamic acid (E) residues, at amino acid residues 415, 452, 489, and 526 are marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of M6_Spy0157. Preferred fragments of M6_Spy0157 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 42

MVSSYMFVRGEKMNNKIFLNKEASFLAHTKRKRRFAVTLVGVFFMLLACAGAIGFGQVAYAADEKTVPSHSSPNP
EFPWYGYDAYGKEYPGYNIWTRYHDLRVNLNGSRSYQVYCFNIQSNYPSQKNSFIKNWFKKIEGNGKSFVDYAHT

TKLGKEELEQRLLSLLYNAYPNDANGYMKGLEHLNAITVTQYAVWHYSDNSQYQFETLWESEAKEGKISRSQVTL
MREALKKLIDPNTEATAVNKTPSGYRLNIFESENEAYQNLLSAEYVPDDPPKPGETSEHNPKTPELDGTPIPEDP
KHPDDNLEPTLPPVMLDGEEVPEVPSESLEPALPPLMPELDGQEVPEKPSIDLPIEVPRYEEFNNKDQSPLAGESG
ETEYITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDT
KEPEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQIETEDTK
EPEVLMGGQSESVEFTKDTQTGMSGFSETATVVEDTRPKLVFHFDNNEPKVEENREKPTKNITPILPATGDIENV
LAFLGILILSVLSIFSLLKNKQSNKKV

M6_Spy0158: M6_Spy0158 is a reverse transcriptase. An example of Spy0158 is shown in the amino acid sequence SEQ ID NO 43.

SEQ ID NO: 43
MSLRHQNKKGIRKEGWKSRPQSRWSDHCQLVAQKSVLKQAISKTVLAERGLFSCLDDYLERHALKVN

M6_Spy0159: M6_Spy0159 is a collagen adhesion protein. It contains a sortase substrate motif LPXSG, shown in italics in the amino acid sequence SEQ ID NO: 44.

SEQ ID NO: 44
MYSRLKRELVIVINRKKKYKLIRLMVTVGLIFSQLVLPIRRLGLQMISTQTKVIPQEIVTQTETQGTQVVATKQK
LESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAIVTMALELRKQGLSIVDTKIVRIQSSTNQRN
DITTTLTFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIVSTTRLKE
EQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNILKPKLSSIESSLDSKDFEIVKTIDNLYTWDDQFYLLDFI
SKQYEVLKTDYQSAKDSTPQTRDILFGEYTVEPLVMNKGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRSLT
PRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKEDEHDTSDLYRLYLDMTGKKNPLDILVVVDKSG
SMQEGIGSVQRYRYYAQRWDDYYSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVN
GLLQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN
GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQEGSKLAIDEFKARY
PNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDK
QPDVLVTRKSKVNDETEILYQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKAS
DEAYEKYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
ADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYYLYETKAKLGYTLPENPWEVAVANNGDIK
VKHPIEGELKSKDGSYMIKNYKIYQ*LPSSG*GRGSQIFIIVGSMTATVALLFYRRQHRKKQY

M6_Spy0159 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 181** LPSSG (shown in italics in SEQ ID NO: 44, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant M6_Spy0159 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in M6_Spy0159. The pilin motif sequence is underlined in SEQ ID NO: 44, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 265 and 276. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of M6_Spy0159 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 44
MYSRLKRELVIVINRKKKYKLIRLMVTVGLIFSQLVLPIRRLGLQMISTQTKVIPQEIVTQTETQGTQVVATKQK
LESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAIVTMALELRKQGLSIVDTKIVRIQSSTNQRN
DITTTLTFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIVSTTRLKE
EQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNILKPKLSSIESSLDSKDFEIVKTIDNLYTWDDQFYLLDFI
SKQYEVLKTDYQSAKDSTPQTRDILFGEYTVEPLVMNKGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRSLT
PRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKEDEHDTSDLYRLYLDMTGKKNPLDILVVVDKSG

SMQEGIGSVQRYRYYAQRWDDYYSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVN
GLLQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN
GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQEGSKLAIDEFKARY
PNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDK
QPDVLVTRKSKVNDETEILYQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKAS
DEAYEKYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
ADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYYLYETKAKLGYTLPENPWEVAVANNGDIK
VKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTATVALLFYRRQHRKKQY

An E box containing a conserved glutamic residue has been identified in M6_Spy0159. The E-box motif is underlined in SEQ ID NO: 44, below. The conserved glutamic acid (E), at amino acid residue 950, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of M6_Spy0159. Preferred fragments of M6_Spy0159 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 44

MYSRLKRELVIVINRKKKYKLIRLMVTVGLIFSQLVLPIRRLGLQMISTQTKVIPQEIVTQTETQGTQVVATKQK
LESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAIVTMALELRKQGLSIVDTKIVRIQSSTNQRN
DITTTLTFKNGLSLEGASTEANDPNVRVGIVNPNDTVQTITPTIKQDADGKVKNLVFTGRLGKQVIIVSTTRLKE
EQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNILKPKLSSIESSLDSKDFEIVKTIDNLYTWDDQFYLLDFI
SKQYEVLKTDYQSAKDSTPQTRDILFGEYTVEPLVMNKGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRSLT
PRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKEDEHDTSDLYRLYLDMTGKKNPLDILVVVDKSG
SMQEGIGSVQRYRYYAQRWDDYYSQWVYHGTFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVN
GLLQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN
GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQEGSKLAIDEFKARY
PNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDK
QPDVLVTRKSKVNDETEILYQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKAS
DEAYEKYKDNEGRYSEMGDSDTDYGTNQTSSGKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
ADNNQKKLAGVEFELRKEDKKIVWEKGTTGSNGQLNFKYLQKGKTYYLYETKAKLGYTLPENPWEVAVANNGDIK
VKHPIEGELKSKDGSYMIKNYKIYQLPSSGGRGSQIFIIVGSMTATVALLFYRRQHRKKQY

M6_Spy0160: M6_Spy0160 is a fimbrial structural subunit. It contains a sortase substrate motif LPXTG (SEQ ID NO: 122), shown in italics in amino acid sequence SEQ ID NO: 45.

SEQ ID NO: 45

MTNRRETVREKILITAKKLMLACLAILAVVGLGMTRVSALSKDDTAQLKITNIEGGPTVTLYKIGEGVYNTNGDS
FINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENVSISNGTATYNARGASVYIALLTGATDGRTYNPI
LLAASYNGEGNLVTKNIDSKSNYLYGQTSVAKSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFELPSYTKEA
VNKTVYVSDNMSEGLTFNFNSLTVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYK
AVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDSKIVYTYQIAFRKVDSVSKTPLIGA
IFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTS
DKNKATDNSEQVGWLKNGIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGE*LPS*
*TG*SIGTYLFKAIGSAAMIGAIGIYIVKRRKA

M6_Spy0160 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 131** LPSTG (shown in italics in SEQ ID NO: 45, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant M6_Spy0160 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

An E box containing a conserved glutamic residue has been identified in M6_Spy0160. The E-box motif is underlined in SEQ ID NO: 45, below. The conserved glutamic acid (E), at amino acid residue 412, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of M6_Spy0160. Preferred fragments of M6_Spy0160 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 45

MTNRRETVREKILITAKKLMLACLAILAVVGLGMTRVSALSKDDTAQLKITNIEGGPTVTLYKIGEGVYNTNGDS
FINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENVSISNGTATYNARGASVYIALLTGATDGRTYNPI
LLAASYNGEGNLVTKNIDSKSNYLYGQTSVAKSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFELPSYTKEA
VNKTVYVSDNMSEGLTFNFNSLTVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYK
AVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDSKIVYTYQIAFRKVDSVSKTPLIGA
IFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTS
DKNKATDNSEQVGWLKNGIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDIPNTKLGELPS
TGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA

M6_Spy0161 is a srtB type sortase. An example of an amino acid sequence of M6_Spy-161 is shown in SEQ ID NO: 46.

SEQ ID NO: 46

MTERLKNLGILLLFLLGTAIFLYPTLSSQWNAYRDRQLLSTYHKQVIQKKPSEMEEVWQKAKAYNARLGIQPVPD
AFSFRDGIHDKNYESLLQIENNDIMGYVEVPSIKVTLPIYHYTTDEVLTKGAGHLFGSALPVGGDGTHTVISAHR
GLPSAEMFTNLNLVKKGDTFYFRVLNKVLAYKVDQILIVEPDQATSLSGVMGKDYATLVTCTPYGVNTKRLLVRG
HRIAYHYKKYQQAKKAMKLVDKSRMWAEVVCAAFGVVIAIILVFMYSRVSAKKSK

As discussed above, applicants have also determined the nucleotide and encoded amino acid sequence of fimbrial structural subunits in several other GAS AI-1 strains of bacteria. Examples of sequences of these fimbrial structural subunits are set forth below.

M6 strain isolate CDC SS 410 is a GAS AI-1 strain of bacteria. CDC SS 410_fimbrial is thought to be a fimbrial structural subunit of M6 strain isolate CDC SS 410. An example of a nucleotide sequence encoding the CDC SS 410_fimbrial protein (SEQ ID NO: 267) and a CDC SS 410_fimbrial protein amino acid sequence (SEQ ID NO: 268) are set forth below.

SEQ ID NO: 267

aaagatgatactgcacaactaaagataacaaatattgaaggtgggccaacagtaacactt
tataaaatagagaaggtgtttacaacactaatggtgattcttttattaactttaaatat
gctgaggggtttctttaactgaaacaggacctacatcacaagaaattactactattgca
aatggtattaatacgggtaaaataaagccttttagtactgaaaacgttagtatttctaat
ggaacagcaacttataatgcgagaggtgcatctgtttatattgcattattaacaggtgcg
acagatggccgtacctacaatcctatttttattagctgcatcttataatggtgagggaaat
ttagttactaaaaatattgattccaaatctaattatttatatggacaaacaagtgttgca
aaatcatcattaccatctattacaaagaaagtaaccgggacaatagatgacgtgaataaa
aagactacctcgttaggaagtgtattgtcttattcgctgacatttgaattaccaagttat
accaaagaagcagtcaataaaacagtatatgtttctgataatatgtcggaaggtcttact
tttaactttaatagtcttacagtagaatggaaaggtaagatggctaatattactgaagat
ggttcagtaatggtagaaaatacaaaaatcggaatagctaaggaggttaataacggtttt
aatttaagttttatttatgatagtttagaatctatatcaccaaatataagttataaagct
gttgtaaacaataaagctattgttggtgaagagggtaatcctaataaagctgaattcttc
tattcaaataatccaacaaaaggtaatacatacgataatttagataagaagcctgataaa
gggaatggtattacatccaaagaagattctaaaattgtttatacttatcaaatagcgttt
agaaaagttgatagtgttagtaagaccccacttattggtgcaatttttggagtttatgat
actagtaataaattaattgatattgttacaaccaataaaaatggatatgctatttcaaca

caagtatcttcaggaaaatataaaaattaaggaattaaaagctcctaaaggttattcattg
aatacagaaacttatgaaattacggcaaattgggtaactgctacagtcaagacaagtgct
aattcaaaaagtactacttatacatctgataaaaataaggcgacagataattcagagcaa
gtaggatggttaaaaaatggtatattctattctatagatagtagacctacaggaaatgat
gttaaagaggcttatattgaatctactaaggctttaactgatggaacaactttctcaaaa
tcgaatgaaggttcaggtacagtattattagaaactgacatccctaacaccaagctaggt
gaactc

SEQ ID NO: 268

KDDTAQLKITNIEGGPTVTLYKIGEGVYNTNGDSFINFKYAEGV
SLTETGPTSQEITTIANGINTGKIKPFSTENVSISNGTATYNARGASVYIALLTGATD
GRTYNPILLAASYNGEGNLVTKNIDSKSNYLYGQTSVAKSSLPSITKKVTGTIDDVNK
KTTSLGSVLSYSLTFELPSYTKEAVNKTVYVSDNMSEGLTFNFNSLTVEWKGKMANIT
EDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNK
AEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDSKIVYTYQIAFRKVDSVSKTPLIGA
IFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGYSLNTETYEITANWV
TATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIFYSIDSRPTGNDVKEAYIESTK
ALTDGTTFSKSNEGSGTVLLETDIPNTKLGEL

M6 strain isolate ISS 3650 is a GAS AI-1 strain of bacteria. ISS3650_fimbrial is thought to be a fimbrial structural subunit of M6 strain isolate ISS 3650. An example of a nucleotide sequence encoding the ISS3650_fimbrial protein (SEQ ID NO: 269) and an ISS3650_fimbrial protein amino acid sequence (SEQ ID NO: 270) are set forth below.

SEQ ID NO: 269

gaatggaaaggtaagatggctaatattactgaagatggttcagtaatggtagaaaataca
aaaatcggaatagctaaggaggttaataacggttttaatttaagttttatttatgatagt
ttagaatctatatcaccaaatataagttataaagctgttgtaaacaataaagctattgtt
ggtgaagagggtaatcctaataaagctgaattcttctattcaaataatccaacaaaaggt
aatacatacgataatttagataagaagcctgataaagggaatggtattacatccaaagaa
gattctaaaattgtttatacttatcaaatagcgtttagaaaagttgatagtgttagtaag
accccacttattggtgcaatttttggagtttatgatactagtaataaattaattgatatt
gttacaaccaataaaaatggatatgctatttcaacacaagtatcttcaggaaaatataaa
attaaggaattaaaagctcctaaaggttattcattgaatacagaaacttatgaaattacg
gcaaattgggtaactgctacagtcaagacaagtgctaattcaaaaagtactacttataca
tctgataaaaataaggcgacagataattcagagcaagtaggatggttaaaaaatggtata
ttctattctatagatagtagacctacaggaaatgatgttaaagaggcttatattgaatct
actaaggctttaactgatggaacaactttctcaaaatcgaatgaaggttcaggtacagta
ttattagaaactgacatcc

SEQ ID NO: 270

EWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESI
SPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPTKGNTYDNLDKKPDKGNGITSKEDS
KIVYTYQIAFRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYK
IKELKAPKGYSLNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKN
GIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDI

M23 strain isolate DSM2071 is a GAS AI-1 strain of bacteria. DSM2071_fimbrial is thought to be a fimbrial structural subunit of M23 strain DSM2071. An example of a nucleotide sequence encoding the DSM2071_fimbrial protein (SEQ ID NO: 251) and a DSM2071_fimbrial protein amino acid sequence (SEQ ID NO: 252) are set forth below.

SEQ ID NO: 251

atgagagagaaaatattaatagcagcaaaaaaactaatgctagcttgtttagctatctta
gctgtagtagggcttggaatgacaagagtatcagctttatcaaaagatgataaggcggag
ttgaagataacaaatatcgaaggtaaaccgaccgtgacactgtataaaattggtgatgga
aaatacagtgagcgaggggattcttttattggatttgagttaaagcaaggtgtggagcta
aataaggcaaaacctacatctcaagaaataaataaaatcgctaatggtattaataaaggt
agtgttaaggctgaagtagttaatataaaagaacatgctagtacaacttatagttataca

agaacttgtgcatgtattacttcgggtaaattgactggagctactgatggacgtgcctat
aatcctatcttactgacagcttcttacaatgaggaaaatccacttaagggagggcagatt
gacgcaactagtcattatctttttggagaagaagcagttgctaaatctagccaaccaaca
attagcaagtcaattacaaaatccacaaaagatggtgataaagatacagcatctgtaggt
gaaaaagttgattacaaattaactgttcagttaccaagttattcgaaagatgctatcaat
aaaacggtgtttatcactgacaaattgtctcagggacttactttccttccaaaaagttta
aagattatctggaatggtcaaacgttaacaaaggtgaatgaagaatttaaagctggagat
aaggtaattgctcaacttaaggttgaaaataatggatttaatctgaactttaattatgat
aaccttgataatcatgccccagaagttaactatagtgctctactaaatgaaaacgcagtt
gttggtaaaggtggtaatgacaataatgtagactattactattcaaataatccgaataaa
ggagagacccataaaacaactgagaagcctaaagagggtgaaggtactggtatcactaaa
aagacggataaaaaaaccgtctacacctatcgtgtagcctttaagaaaacaggcaaagat
catgccccactagctggtgctgttttcggtatctattcagataaggaagcgaaacaatta
gtcgatattgttgtgacaaatgcacagggttatgcagcatcaagcgaagttgggaaaggg
acttattacattaaagaaattaaatcccctaagggttactctttaaatacaaatatttat
gaagtggaaacttcatgggaaaaagctacaacgacttctacaactaatcgtttagagaca
atttatacaacagatgataatcaaaagtctccaggaactaatacagttggttggttggaa
gatggtgtcttttacaaagaaaatccaggtggtgatgctaaacttgcctatatcaaacaa
tcaacagaggagacttctacaactatagaagtcaaagaaaatcaagctgaaggttcaggt
acggtattattagaaactgaaattcctaacaccaaattaggtgaattaccttcgacaggt
agcattggtacttacctctttaaagctattggttcggctgctatgatcggtgcaattggt
atttatattgttaaacgtcgtaaagcttaa

SEQ ID NO: 252

MREKILIAAKKLMLACLAILAVVGLGMTRVSALSKDDKAELKIT
NIEGKPTVTLYKIGDGKYSERGDSFIGFELKQGVELNKAKPTSQEINKIANGINKGSV
KAEVVNIKEHASTTYSYTTTGAGIYLAILTGATDGRAYNPILLTASYNEENPLKGGQI
DATSHYLFGEEAVAKSSQPTISKSITKSTKDGDKDTASVGEKVDYKLTVQLPSYSKDA
INKTVFITDKLSQGLTFLPKSLKIIWNGQTLTKVNEEFKAGDKVIAQLKVENNGFNLN
FNYDNLDNHAPEVNYSALLNENAVVGKGGNDNNVDYYYSNNPNKGETHKTTEKPKEGE
GTGITKKTDKKTVYTYRVAFKKTGKDHAPLAGAVFGIYSDKEAKQLVDIVVTNAQGYA
ASSEVGKGTYYIKEIKSPKGYSLNTNIYEVETSWEKATTTSTTNRLETIYTTDDNQKS
PGTNTVGWLEDGVFYKENPGGDAKLAYIKQSTEETSTTIEVKENQAEGSGTVLLETEI
PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA

GAS AI-2 sequences

As discussed above, a GAS AI-2 sequence is present in an M1 strain isolate (SF370). Examples of GAS AI-2 sequences from M1 strain isolate SF370 are set forth below.

Spy0124 is a rofA transcriptional regulator. An example of an amino acid sequence for Spy0124 is set forth in SEQ ID NO:47.

SEQ ID NO: 47

MIEKYLESSIESKCQLIVLFFKTSYLPITEVAEKTGLTFLQLNHYCEELNAFFPGSLSMTIQKRMISCQFTHPFK
ETYLYQLYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNSSAYRMREALIPLLRNFELKLSKNKIVGEEYRIR
YLIALLYSKFGIKVYDLTQQDKNTIHSFLSHSSTHLKTSPWLSESFSFYDILLALSWKRHQFSVTIPQTRIFQQL
KKLFVYDSLKKSSHDIIETYCQLNFSAGDLDYLYLIYITANNSFASLQWTPEHIRQYCQLFEENDTFRLLLNPII
TLLPNLKEQKASLVKALMFFSKSFLFNLQHFIPETNLFVSPYYKGNQKLYTSLKLIVEEWMAKLPGKRDLNHKHF
HLFCHYVEQSLRNIQPPLVVVFVASNFINAHLLTDSFPRYFSDKSIDFHSYYLLQDNVYQIPDLKPDLVITHSQL
IPFVHHELTKGIAVAEISFDESILSIQELMYQVKEEKFQADLTKQLT

GAS 015 is also referred to as Cpa. It contains a sortase substrate motif VVXTG (SEQ ID NO: 135), shown in italics in SEQ ID NO: 48.

SEQ ID NO: 48

LRGEKMKKTRFPNKLNTLNTQRVLSKNSKRFTVTLVGVFLMIFALVTSMVGAKTVFGLVESSTPNAINPDSSSEY
RWYGYESYVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAFPLGSDSSVKKWYKKHDGISTKFEDYAMSPR
ITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQEAVWYYSDNAPISNPDESFKRESESNLVSTSQLSLM
RQALKQLIDPNLATKMPKQVPDDFQLSIFESEDKGDKYNKGYQNLLSGGLVPTKPPTPGDPPMPPNQPQTTSVLI
RKYAIGDYSKLLEGATLQLTGDNVNSFQARVFSSNDIGERIELSDGTYTLTELNSPAGYSIAEPITFKVEAGKVY

TIIDGKQIENPNKEIVEPYSVEAYNDFEEFSVLTTQNYAKFYYAKNKNGSSQVVYCFNADLKSPPDSEDGGKTMT
PDFTTGEVKYTHIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYREKGQAIEYSGLTETQLRAATQLAIYYFTD
SAELDKDKLKDYHGFGDMNDSTLAVAKILVEYAQDSNPPQLTDLDFFIPNNNKYQSLIGTQWHPEDLVDIIRMED
KKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNNKQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLQGL
PEGYSYLVKETDSEGYKVKVNSQEVANATVSKTGITSDETLAFENNKEP*VVPTG*VDQKINGYLALIVIAGISLGI
WGIHTIRIRKHD

GAS 015 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 182** VVPTG (shown in italics in SEQ ID NO: 48, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant GAS 015 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in GAS 015. The pilin motif sequence is underlined in SEQ ID NO: 48, below. Conserved lysine (K) residues are also marked in bold, at amino acid residue 243. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of GAS 015 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 48

LRGEKMKKTRFPNKLNTLNTQRVLSKNSKRFTVTLVGVFLMIFALVTSMVGAKTVFGLVESSTPNAINPDSSSEY
RWYGYESYVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAFPLGSDSSVKKWYKKHDGISTKFEDYAMSPR
ITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQEAVWYYSDNAPISNPDESFKRESESNLVSTSQLSLM
RQALKQLIDPNLATKMPKQVPDDFQLSIFESEDKGDKYNKGYQNLLSGGLVPTKPPTPGDPPMPPNQPQTTSVLI
RKYAIGDYSKLLEGATLQLTGDNVNSFQARVFSSNDIGERIELSDGTYTLTELNSPAGYSIAEPITFKVEAGKVY
TIIDGKQIENPNKEIVEPYSVEAYNDFEEFSVLTTQNYAKFYYAKNKNGSSQVVYCFNADLKSPPDSEDGGKTMT
PDFTTGEVKYTHIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYREKGQAIEYSGLTETQLRAATQLAIYYFTD
SAELDKDKLKDYHGFGDMNDSTLAVAKILVEYAQDSNPPQLTDLDFFIPNNNKYQSLIGTQWHPEDLVDIIRMED
KKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNNKQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLQGL
PEGYSYLVKETDSEGYKVKVNSQEVANATVSKTGITSDETLAFENNKEPVVPTGVDQKINGYLALIVIAGISLGI
WGIHTIRIRKHD

An E box containing a conserved glutamic residue has been identified in GAS 015. The E-box motif is underlined in SEQ ID NO: 48, below. The conserved glutamic acid (E), at amino acid residue 352, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of GAS 015. Preferred fragments of GAS 015 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 48

LRGEKMKKTRFPNKLNTLNTQRVLSKNSKRFTVTLVGVFLMIFALVTSMVGAKTVFGLVESSTPNAINPDSSSEY
RWYGYESYVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAFPLGSDSSVKKWYKKHDGISTKFEDYAMSPR
ITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQEAVWYYSDNAPISNPDESFKRESESNLVSTSQLSLM
RQALKQLIDPNLATKMPKQVPDDFQLSIFESEDKGDKYNKGYQNLLSGGLVPTKPPTPGDPPMPPNQPQTTSVLI
RKYAIGDYSKLLEGATLQLTGDNVNSFQARVFSSNDIGERIELSDGTYTLTELNSPAGYSIAEPITFKVEAGKVY
TIIDGKQIENPNKEIVEPYSVEAYNDFEEFSVLTTQNYAKFYYAKNKNGSSQVVYCFNADLKSPPDSEDGGKTMT
PDFTTGEVKYTHIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYREKGQAIEYSGLTETQLRAATQLAIYYFTD
SAELDKDKLKDYHGFGDMNDSTLAVAKILVEYAQDSNPPQLTDLDFFIPNNNKYQSLIGTQWHPEDLVDIIRMED
KKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNNKQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLQGL

PIGVSYTVKETDSEGYKVEKVNGQEVANATVGKTGITSDETLAFENNKEPVVPTGVDQKINGYLALIVIAGISLGI
WGIHTIRIRKHD

Spy0127 is a LepA putative signal peptidase. An example of an amino acid sequence for Spy0127 is set forth in SEQ ID NO: 49.

SEQ ID NO: 49
MIIKRNDMAPSVKAGDAILFYRLSQTYKVEEAVVYEDSKTSITKVGRIIAQAGDEVDLTEQGELKINGHIQNEGL
TFIKSREANYPYRIADNSYLILNDYYSQESENYLQDAIAKDAIKGTINTLIRLRNH

Spy0128 is thought to be a fibrial protein. It contains a sortase substrate motif EVXTG (SEQ ID NO: 136) shown in italics in SEQ ID NO: 50.

SEQ ID NO: 50
MKLRHLLLTGAALTSFAATTVHGETVVNGAKLTVTKNLDLVNSNALIPNTDFTFKIEPDTTVNEDGNKFKGVALN
TPMTKVTYTNSDKGGSNTKTAEFDFSEVTFEKPGVYYYKVTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQKPVA
TYIVGYKEGSKVPIQFKNSLDSTTLTVKKKVSGTGGDRSKDFNFGLTLKANQYYKASEKVMIEKTTKGGQAPVQT
EASIDQLYHFTLKDGESIKVTNLPVGVDYVVTEDDYKSEKYTTNVEVSPQDGAVKNIAGNSTEQETSTDKDMTIT
FTNKKDF*EVPTG*VAMTVAPYIALGIVAVGGALYFVKKKNA

Spy0128 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 183** EVPTG (shown in italics in SEQ ID NO: 50, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Spy0128 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two E boxes containing a conserved glutamic residue have been identified in Spy0128. The E-box motifs are underlined in SEQ ID NO: 50, below. The conserved glutamic acid (E) residues, at amino acid residues 271 and 290, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of Spy0128. Preferred fragments of Spy0128 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 50
MKLRHLLLTGAALTSFAATTVHGETVVNGAKLTVTKNLDLVNSNALIPNTDFTFKIEPDTTVNEDGNKFKGVALN
TPMTKVTYTNSDKGGSNTKTAEFDFSEVTFEKPGVYYYKVTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQKPVA
TYIVGYKEGSKVPIQFKNSLDSTTLTVKKKVSGTGGDRSKDFNFGLTLKANQYYKASEKVMIEKTTKGGQAPVQT
EASIDQLYHFTLKDGESIKVTNLPVGVDYVVTEDDYKSEKYT**T**NVEVSPQDGAVKNIAGN**S**TEQETSTDKDMTIT
FTNKKDFEVPTGVAMTVAPYIALGIVAVGGALYFVKKKNA

Spy0129 is a srtC1 type sortase. An example of an amino acid sequence for Spy0129 is set forth in SEQ ID NO: 51.

SEQ ID NO: 51
MIVRLIKLLDKLINVIVLCFFFLCLLIAALGIYDALTVYQGANATNYQQYKKKGVQFDDLLAINSDVMAWLTVKG
THIDYPIVQGENNLEYINKSVEGEYSLSGSVFLDYRNKVTFEDKYSLIYAHHMAGNVMFGELPNFRKKSFFNKHK
EFSIETKTKQKLKINIFACIQTDAFDSLLFNPIDVDISSKNEFLNHIKQKSVQYREILTTNESRFVALSTCEDMT
TDGRIIVIGQIE"

Spy0130 is referred to as a hypothetical protein. It contains a sortase substrate motif LPXTG (SEQ ID NO: 122), shown in italics in SEQ ID NO: 52.

SEQ ID NO: 52
MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVTT
SVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVVIYVLYNEQSGALETNLVSNKLGETEKSELI
FKQEYSEKTPEPHQPDTTEKEKPQKKRNGI*LPSTG*EMVSYVSALGIVLVATITLYSIYKKLKTSK

Spy0130 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 131** LPSTG (shown in italics in SEQ ID NO: 52, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Spy0130 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two E boxes containing conserved glutamic residues have been identified in Spy0130. The E-box motifs are underlined in SEQ ID NO: 52, below. The conserved glutamic acid (E) residues, at amino acid residues 118 and 148, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of Spy0130. Preferred fragments of Spy0130 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 52
MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVTT
SVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVY**E**VVIYVLYNEQSGALETNLVSNKLGETEKSELI
FKQEYSEKTPEPHQPDTTEKEKPQKKRNGILPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK

Spy0131 is referred to as a conserved hypothetical protein. An example of an amino acid sequence of Spy0131 is set forth in SEQ ID NO: 53

SEQ ID NO: 53
MTRTNYQKKRMTCPVETEDITYRRKKIKGRRQAILAQFEPELVHHELIGDSCTCPDCHGTLTEIGSVVQRQELVF
IPAQLKRINHVQHAYKCQTCSDNSLSDKIIKAPVPKAPLAHSLGSASIIAHTVHQKFTLKVPNYRQEEDWNKLGL
SISRKEIANWHIKSSQYYFEPLYDLLRDILLSQEVIHADETSYRVLESDTQLTYYWTFLSGKHEKKGITLYHHDK
RRSGLVTQEVLGDYSGYVHCDMHGAYRQLEHAKLVGCWAHVRRKFFEATPKQADKTSLGRKGLVYCDKLFALEAE
WCELPPQERLVKRKEILTPLMTTFFDWCREQVVLSGSKLGLAIAYSLKHERTFRTVLEDGHIVLSNNMAERAIKS
LVMGRKNWLFSQSFEGAKAAAIIMSLLETAKRHGLNSEKYISYLLDRLPNEETLAKREVLEAYLPWAKKVQTNCQ

Spy0133 is referred to as a conserved hypothetical protein. An example of an amino acid sequence of Spy0133 is set forth in SEQ ID NO: 54.

SEQ ID NO: 54
MTIRLNDLGQVYLVCGKTDMRQGIDSLAYLVKSQHELDLFSGAVYLFCGGRRDRFKALYWDGQGFWLLYKRFENG
KLAWPRNRDEVKCLTAVQVDWLMKGFFISPNIKISKSHDFY

Spy0135 is a SrtB type sortase. It is also referred to as a putative fibria-associated protein. An example of an amino acid sequence of Spy0135 is set forth in SEQ ID NO: 55.

SEQ ID NO: 55
MECYRDRQLLSTYHKQVTQKKPSEMEEVWQKAKAYNARLGIQPVPDAFSFRDGIHDKNYESLLQIENNDIMGYVE
VPSIKVTLPIYHYTTDEVLTKGAGHLFGSALPVGGDGTHTVISAHRGLPSAEMFTNLNLVKKGDTFYFRVLNKVL
AYKVDQILTVEPDQVTSLSGVMGKDYATLVTCTPYGVNTKRLLVRGHRIAYHYKKYQQAKKAMKLVDKSRMWAEV
VCAAFGVVIAIILVFMYSRVSAKKSK

GAS AI-3 sequences

As discussed above, a GAS AI-3 sequence is present in a M3, M18 and M5 strain isolates. Examples of GAS AI-3 sequences from M3 strain isolate MGAS315 are set forth below.

SpyM30097 is as a negative transcriptional regulator (Nra). An example of an amino acid sequence of SpyM30097 is set forth in SEQ ID NO: 56.

SEQ ID NO: 56
MPYVKKKKDSFLVETYLEQSIRDKSELVLLLFKSPTIIFSHVAKQTGLTAVQLKYYCKELDDFFGNNLDITIKKG
KIICCFVKPVKEFYLHQLYDTSTILKLLVFFIKNGTSSQPLIKFSKKYFLSSSSAYRLRESLIKLLREFGLRVSK
NTIVGEEYRIRYLIAMLYSKFGIVIYPLDHLDNQIIYRFLSQSATNLRTSPWLEEPFSFYNMLLALSWKRHQFAV
SIPQTRIFRQLKKLFIYDCLTRSSRQVIENAFSLTFSQGDLEYLFLIYITTNNSFASLQWTPQHIETCCHIFEKN
DTFRLLLEPILKRLPQLNHSKQDLIKALMYFSKSFLFNLQHFVIEIPSFSLPTYTGNSNLYKALKNIVNQWLAQL
PGKRHLNEKHLQLFCSHIEQILKNKQPALTVVLISSNFINAKLLTDTIPRYFSDKGIHFYSFYLLRDDIYQIPSL
KPDLVITHSRLIPFVKNDLVKGVTVAEFSFDNPDYSIASIQNLIYQLKDKKYQDFLNEQLQ

SpyM30098 is thought to be a collagen binding protein (Cpb). It contains a sortase substrate motif VPXTG (SEQ ID NO: 137) shown in italics in SEQ ID NO: 57.

SEQ ID NO: 57
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNKQSSVQDYPWYGYDSYSKGYPD
YSPLKTYHNLKVNLDGSKEYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDGQLQQNILRIL
YNGYPNDRNGIMKGIDPLNAILVTQNAIWYYTDSSYISDTSKAFQQEETDLKLDSQQLQLMRNALKRLINPKEVE
SLPNQVPANYQLSIFQSSDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKYAEGDYSKLLEGATLKLAQI
EGSGFQEKIFDSNKSGEKVELPNGTYVLSELKPPQGYGVATPITFKVAAEKVLIKNKEGQFVENQNKEIAEPYSV
TAFNDFEEIGYLSDFNNYGKFYYAKNTNGTNQVVYCFNADLHSPPDSYDHGANIDPDVSESKEIKYTHVSGYDLY
KYAATPRDKDADFFLKHIKKILDKGYKKKGDTYKTLTEAQFRAATQLAIYYYTDSADLTTLKTYNDNKGYHGFDK
LDDATLAVVHELITYAEDVTLPMTQNLDFFVPNSSRYQALIGTQYHPNELIDVISMEDKQAPIIPITHKLTISKT
VTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETGASDY
EVSVNGKNAPDGKATKASVKEDETVAFENRKDL*VPPTG*LTTDGAIYLWLLLLVPFGLLVWLFGRKGTKK

SpyM30098 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 184** VPPTG (shown in italics in SEQ ID NO: 57, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM30098 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM30098. The pilin motif sequence is underlined in SEQ ID NO: 57, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 262 and 270. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM30098 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 57
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNKQSSVQDYPWYGYDSYSKGYPD
YSPLKTYHNLKVNLDGSKEYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDGQLQQNILRIL
YNGYPNDRNGIMKGIDPLNAILVTQNAIWYYTDSSYISDTSKAFQQEETDLKLDSQQLQLMRNALKRLINPKEVE
SLPNQVPANYQLSIFQSSDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKYAEGDYSKLLEGATLKLAQI
EGSGFQEKIFDSNKSGEKVELPNGTYVLSELKPPQGYGVATPITFKVAAEKVLIKNKEGQFVENQNKEIAEPYSV
TAFNDFEEIGYLSDFNNYGKFYYAKNTNGTNQVVYCFNADLHSPPDSYDHGANIDPDVSESKEIKYTHVSGYDLY
KYAATPRDKDADFFLKHIKKILDKGYKKKGDTYKTLTEAQFRAATQLAIYYYTDSADLTTLKTYNDNKGYHGFDK
LDDATLAVVHELITYAEDVTLPMTQNLDFFVPNSSRYQALIGTQYHPNELIDVISMEDKQAPIIPITHKLTISKT

VTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETGASDY
EVSVNGKNAPDGKATKASVKEDETVAFENRKDLVPPTGLTTDGAIYLWLLLLVPFGLLVWLFGRKGTKK

An E box containing a conserved glutamic residue has been identified in SpyM30098. The E-box motif is underlined in SEQ ID NO: 57, below. The conserved glutamic acid (E), at amino acid residue 330, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyM30098. Preferred fragments of SpyM30098 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 57

MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNKQSSVQDYPWYGYDSYSKGYPD
YSPLKTYHNLKVNLDGSKEYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDGQLQQNILRIL
YNGYPNDRNGIMKGIDPLNAILVTQNAIWYYTDSSYISDTSKAFQQEETDLKLDSQQLQLMRNALKRLINPKEVE
SLPNQVPANYQLSIFQSSDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKYAEGDYSKLLEGATLKLAQI
EGSGFQEKIFDSNKSGEKVELPNGTYVLSELKPPQGYGVATPITFKVAAEKVLIKNKEGQFVENQNKEIAEPYSV
TAFNDFEEIGYLSDFNNYGKFYYAKNTNGTNQVVYCFNADLHSPPDSYDHGANIDPDVSESKEIKYTHVSGYDLY
KYAATPRDKDADFFLKHIKKILDKGYKKKGDTYKTLTEAQFRAATQLAIYYYTDSADLTTLKTYNDNKGYHGFDK
LDDATLAVVHELITYAEDVTLPMTQNLDFFVPNSSRYQALIGTQYHPNELIDVISMEDKQAPIIPITHKLTISKT
VTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETGASDY
EVSVNGKNAPDGKATKASVKEDETVAFENRKDLVPPTGLTTDGAIYLWLLLLVPFGLLVWLFGRKGTKK

SpyM30099 is referred to as LepA. An example of an amino acid sequence of SpyM30099 is set forth in SEQ ID NO: 58.

SEQ ID NO: 58

MTNYLNRLNENPLLKAFIRLVLKISIIGFLGYILFQYVFGVMIVNTNQMSPAVSAGDGVLYYRLTDRYHINDVVV
YEVDDTLKVGRIAAQAGDEVNFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILNDYREERLDSR
YYGALPINQIKGKISTLLRVRGI

SpyM30100 is thought to be a fimbrial protein. An example of an amino acid sequence of SpyM30100 is set forth in SEQ ID NO: 59.

SEQ ID NO: 59

MKKNKLLLATAILATALGTASLNQNVKAETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVEADSTASGKTK
DGLEIKPGIVNGLTEQIISYTNTDKPDSKVKSTEFDFSKVVFPGIGVYRYTVSEKQGDVEGITYDTKKWTVDVYV
GNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQ
KGNEKFEVKIGTPYKFKLKNGESIQLDKLPVGITYKVNEMEANKDGYKTTASLKEGDGQSKMYQLDMEQKTDESA
DEIVVTNKRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyM30100 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 140** QVPTG (shown in italics in SEQ ID NO: 59, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM30100 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, discussed above, containing conserved lysine (K) residues have also been identified in SpyM30100. The pilin motif sequences are underlined in SEQ ID NO: 59, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 57 and 63 and at amino acid residues 161 and 166. The pilin sequences, in particular the conserved lysine residues, are

thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM30100 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 59

MKKNKLLLATAILATALGTASLNQNVKAETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVEADSTASGKTK
DGLEIKPGIVNGLTEQIISYTNTDKPDSKVKSTEFDFSKVVFPGIGVYRYTVSEKQGDVEGITYDTKKWTVDVYV
GNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQ
KGNEKFEVKIGTPYKFKLKNGESIQLDKLPVGITYKVNEMEANKDGYKTTASLKEGDGQSKMYQLDMEQKTDESA
DEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

Two E boxes, each containing a conserved glutamic residue, have been identified in SpyM30100. The E-box motifs are underlined in SEQ ID NO: 59, below. The conserved glutamic acid (E) residues, at amino acid residues 232 and 264, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyM30100. Preferred fragments of SpyM30100 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 59

MKKNKLLLATAILATALGTASLNQNVKAETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVEADSTASGKTK
DGLEIKPGIVNGLTEQIISYTNTDKPDSKVKSTEFDFSKVVFPGIGVYRYTVSEKQGDVEGITYDTKKWTVDVYV
GNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQ
KGNEKFEVKIGTPYKFKLKNGESIQLDKLPVGITYKVNEMEANKDGYKTTASLKEGDGQSKMYQLDMEQKTDESA
DEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyM30101 is a SrtC2 type sortase. An example of an amino acid sequence of SpyM30101 is set forth in SEQ ID NO: 60.

SEQ ID NO: 60

MTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVIGWLNIPG
THIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIPKFLKKDFFSKHN
KAIIETKERKKLTVTIFACLKTDAFNQLVFNPNAITNQDQQRQLVDYISKRSKQFKPVKLKHHTKFVAFSTCENF
STDNRVIVVGTIQE

SpyM30102 is referred to as a hypothetical protein. An example of an amino acid sequence of SpyM30102 is set forth in SEQ ID NO: 61.

SEQ ID NO: 61

MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQY
TYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNIPKTP*L*
*PLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyM30102 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 185** LPLAG (shown in italics in SEQ ID NO: 61, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM30102 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM30102. The pilin motif sequence is underlined in SEQ ID NO: 61, below. The conserved lysine (K) residue is also marked in bold, at amino acid residue 132. The pilin sequence, in

particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM30102 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 61

MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQY
TYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNIPKTPL
PLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

Two E boxes containing conserved glutamic residues have been identified in SpyM30102. The E-box motifs are underlined in SEQ ID NO: 61, below. The conserved glutamic acid (E) residues, at amino acid residues 52 and 122, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyM30102. Preferred fragments of SpyM30102 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 61

MILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTFTTVGQY
TYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNIPKTPL
PLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyM30103 is referred to as a putative multiple sugar metabolism regulator. An example of an amino acid sequence for SpyM3103 is set forth in SEQ ID NO: 62.

SEQ ID NO: 62

MVRFDLKHVQTLHSLSQLPISVMSQDKALIQVYGNDDYLLCYYQFLKHLAIPQAAQDVIFYEGLFEESFMIFPLC
HYIIAIGPFYPYSLNKDYQEQLANNCLKHSSHRSKEELLSYMALVPHFPINNVRNLLIAIDAFFDTQFETTCQQT
IHQLLQHSKQMTADPDIIHRLKHISKASSQLPPVLEHLNHIMDLVKLGNPQLLKQEINRIPLSSITSSSISALRA
EKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEASDLLKVLRIRCAAIIHFSESLTNKSISDKRQMYNS
VLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSLQHYILSTKIKEAQLLLKRGIPVGEVAKSLYFYDT
THFHKIFKKYTGISSKDYLAKYRDNI

SpyM30104 is thought to be a F2 like fibronectic binding protein. An example of an amino acid sequence for SpyM30104 is set forth in SEQ ID NO: 63.

SEQ ID NO: 63

MSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVLTEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPAD
RSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDLFVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKI
WVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQINSEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLE
PKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHIDITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEF
GKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSS
GKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGS
GQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTEIEDSKSSDVIIGGQGE
VVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNKEP
ESNSEIPKKDKSKSNTS*LPATG*EKQHNKFFWMVTSCSLISSVFVISLKSKKRLSSC

SpyM30104 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 180** LPATG (shown in italics in SEQ ID NO: 63, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM30104 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, discussed above, containing conserved lysine (K) residues have also been identified in SpyM30104. The pilin motif sequences are underlined in SEQ ID NO: 63, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 156 and 227. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM30104 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 63

MSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVLTEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPAD
RSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDLFVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKI
WVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQINSEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLE
PKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHIDITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEF
GKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSS
GKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGS
GQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTEIEDSKSSDVIIGGQGE
VVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNKEP
ESNSEIPKKDKSKSNTSLPATGEKQHNKFFWMVTSCSLISSVFVISLKSKKRLSSC

An E box containing a conserved glutamic residue has been identified in SpyM30104. The E-box motif is underlined in SEQ ID NO: 63, below. The conserved glutamic acid (E), at amino acid residue 402, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyM30104. Preferred fragments of SpyM30104 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 63

MSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVLTEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPAD
RSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDLFVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKI
WVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQINSEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLE
PKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHIDITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEF
GKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSS
GKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGS
GQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTEIEDSKSSDVIIGGQGE
VVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNKEP
ESNSEIPKKDKSKSNTSLPATGEKQHNKFFWMVTSCSLISSVFVISLKSKKRLSSC

Examples of GAS AI-3 sequences from M3 strain isolate SSI-1 are set forth below.

Sps0099 is a negative transcriptional regulator (Nra). An example of an amino acid sequence for Sps0099 is set forth in SEQ ID NO: 64.

SEQ ID NO: 64

MPYVKKKKDSFLVETYLEQSIRDKSELVLLLFKSPTIIFSHVAKQTGLTAVQLKYYCKELDDFFGNNLDITIKKG
KIICCFVKPVKEFYLHQLYDTSTILKLLVFFIKNGTSSQPLIKFSKKYFLSSSSAYRLRESLIKLLREFGLRVSK
NTIVGEEYRIRYLIAMLYSKFGIVIYPLDHLDNQIIYRFLSQSATNLRTSPWLEEPFSFYNMLLALSWKRHQFAV
SIPQTRIFRQLKKLFIYDCLTRSSRQVIENAFSLTFSQGDLEYLFLIYITTNNSFASLQWTPQHIETCCHIFEKN
DTFRLLLEPILKRLPQLNHSKQDLIKALMYFSKSFLFNLQHFVIEIPSFSLPTYTGNSNLYKALKNIVNQWLAQL
PGKRHLNEKHLQLFCSHIEQILKNKQPALTVVLISSNFINAKLLTDTIPRYFSDKGIHFYSFYLLRDDIYQIPSL
KPDLVITHSRLIPFVKNDLVKGVTVAEFSFDNPDYSIASIQNLIYQLKDKKYQDFLNEQLQ

Sps0100 is thought to be a collagen binding protein (Cbp). It contains a sortase substrate motif VPXTG shown in italics in SEQ ID NO: 65.

SEQ ID NO: 65

MQKRDKTNYGSANNKRRQTTIGLLKVELTFVALIGIVGFSIRAFGAEEQSVPNKQSSVQDYPWYGYDSYSKGYPD
YSPLKTYHNLKVNLDGSKEYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDGQLQQNILRIL
YNGYPNDRNGIMKGIDPLNAILVTQNAIWYYTDSSYISDTSKAFQQEETDLKLDSQQLQLMRNALKRLINPKEVE
SLPNQVPANYQLSIFQSSDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKYAEGDYSKLLEGATLKLAQI
EGSGFQEKIFDSNKSGEKVELPNGTYVLSELKPPQGYGVATPITFKVAAEKVLIKNKEGQFVENQNKEIAEPYSV
TAFNDFEEIGYLSDFNNYGKFYYAKNTNGTNQVVYCFNADLHSPPDSYDHGANIDPDVSESKEIKYTHVSGYDLY
KYAATPRDKDADFFLKHIKKILDKGYKKKGDTYKTLTEAQFRAATQLAIYYYTDSADLTTLKTYNDNKGYHGFDK
LDDATLAVVHELITYAEDVTLPMTQNLDFFVPNSSRYQALIGTQYHPNELIDVISMEDKQAPIIPITHKLTISKT
VTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETGASDY
EVSVNGKNAPDGKATKASVKEDETVAFENRKDL*VPPTG*LTTDGAIYLWLLLLVPFGLLVWLFGRKGTKK

Sps0101 is referred to as a LepA protein. An example of an amino acid sequence of Sps0101 is set forth as SEQ ID NO: 66

SEQ ID NO: 66

MTNYLNRLNENPLLKAFIRLVLKISIIGFLGYILFQYVFGVMIVNTNQMSPAVSAGDGVLYYRLTDRYHINDVVV
YEVDDTLKVGRIAAQAGDEVNFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILNDYREERLDSR
YYGALPINQIKGKISTLLRVRGI

Sps0102 is thought to be a fimbrial protein. It contains a sortase substrate motif QVXTG shown in italics in SEQ ID NO: 67.

SEQ ID NO: 67

MEREKMKKNKLLLATAILATALGTASLNQNVKAETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVEADSTA
SGKTKDGLEIKPGIVNGLTEQIISYTNTDKPDSKVKSTEFDFSKVVFPGIGVYRYTVSEKQGDVEGITYDTKKWT
VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLKVKKNVSGNTGELQKEFDFTLTLNESTNFKKDQ
IVSLQKGNEKFEVKIGTPYKFKLKNGESIQLDKLPVGITYKVNEMEANKDGYKTTASLKEGDGQSKMYQLDMEQK
TDESADEIVVTNKRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

Sps0103 is a SrtC2 type sortase. An example of Sps0103 is set forth in SEQ ID NO: 68.

SEQ ID NO: 68

MVMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVIGWLNI
PGTHIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIPKFLKKDFFSK
HNKAIIETKERKKLTVTIFACLKTDAFNQLVFNPNAITNQDQQRQLVDYISKRSKQFKPVKLKHHTKFVAFSTCE
NFSTDNRVIVVGTIQE

Sps0104 is referred to as a hypothetical protein. It contains a sortase substrate motif LPXAG shown in italics in SEQ ID NO: 69.

SEQ ID NO: 69

MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTF
TTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPN
IPKTP*LPLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSRL

Sps0105 is referred to as a putative multiple sugar metabolism regulator. An example of Sps0105 is set forth in SEQ ID NO: 70.

SEQ ID NO: 70

MALVPHFPINNVRNLLIAIDAFFDTQFETTCQQTIHQLLQHSKQMTADPDIIHRLKHISKASSQLPPVLEHLNHI
MDLVKLGNPQLLKQEINRIPLSSITSSSISALRAEKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEAS
DLLKVLRIRCAAIIHFSESLTNKSISDKRQMYNSVLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSL
QHYILSTKIKEAQLLLKRGIPVGEVAKSLYFYDTTHFHKIFKKYTGISSKDYLAKYRDNI

Sps0106 is thought to be a F2 like fibronectic binding protein. It contains a sortase substrate LPXTG (SEQ ID NO: 122) shown in italics in SEQ ID NO: 71.

SEQ ID NO: 71

MTQKNSYKLSFLLSLTGETLGLELVFTSLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEV
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GEVVDTTEDTQSGMTGHSGSTTKIEDSKSSDVIVGGQGQIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNK
EPESNSEIPKKDKSKSNTS*LPATGE*KQHNKFFWMVTSCSLISSVFVISLKSKKRLSSC

Examples of GAS AI-3 sequences from M5 isolate Manfredo are set forth below.

Orf 77 encodes a negative transcription regulator (Nra). An example of the nucleotide sequence encoding Nra (SEQ ID NO: 88) and an Nra amino acid sequence (SEQ ID NO: 89) are set forth below.

SEQ ID NO: 88

ATGCCTTATGTCAAAAAGAAAAAGGATAGTTTCTTAGTAGAAACATATCTTGAACAGTCTATTAGAGATAAAAGT
GAATTAGTCTTACTGTTATTTAAATCGCCTACTATCATTTTTTCTCATGTTGCTAAACAAACTGGTCTGACGGCT
GTACAATTAAAATATTACTGTAAAGAACTTGATGACTTTTTTGGAAATAATTTAGACATTACCATTAAAAAGGGC
AAAATAATATGTTGTTTTGTCAAACCTGTTAAGGAATTCTACCTTCATCAACTCTATGACACATCAACAATATTA
AAATTATTAGTTTTCTTTATTAAAAATGGAACGTCATCACAACCTCTGATTAAATTTTCAAAAAAGTATTTTCTA
TCAAGCTCCTCAGCTTATCGACTACGGGAATCGCTGATCAAATTACTACGGGAATTTGGCTTGAGAGTCTCAAAA
AATACAATTGTCGGAGAGGAATATCGTATTCGCTATCTTATTGCCATGCTATATAGTAAATTTGGCATTGTCATC
TATCCGTTAGATCATCTAGACAATCAAATTATTTATCGCTTCTTATCACAAAGTGCAACCAATTTAAGAACATCG
CCCTGGCTAGAGGAACCTTTTTCTTTTTATAATATGTTACTTGCCTTGTCATGGAAACGTCACCAATTTGCAGTT
AGCATTCCTCAAACACGTATTTTTCGACAATTAAAAAAGCTTTTTATCTATGATTGTTTAACTCGAAGCAGTCGA
CAAGTAATCGAAAATGCTTTTTTCGTTAATGTTCTCACAAGGAGATCTCGATTATCTTTTTTTAATTTATATTACC
ACCAATAATTCCTTTGCCAGCCTACAATGGACTCCACAGCATATTGAAACTTGCTGCCATATTTTTGAAAAAAAT
GACACATTTCGGTTATTGTTAGAGCCCATTCTTAAACGTTTACCGCAATTAAACCATTCTAAACAAGACCTTATT
AAAGCCCTTATGTATTTTTCAAAATCTTTTCTATTTAACCTCCAACATTTCGTCATCGAGATTCCTTCTTTTTCC
TTGCCGACCTATACAGGCAACTCTAATCTTTACAAAGCTTTAAAAAATATTGTAAATCAGTGGCTTGCTCAATTA
CCCGGAAAGCGTCATCTTAACGAAAAGCATCTCCAACTTTTTTGCTCTCATATTGAACAAATCTTAAAAAATAAA
CAACCTGCTTTAACTGTCGTTTTAATATCTAGTAACTTTATAAATGCTAAACTCCTTACAGATACTATCCCACGA
TATTTTTCTGATAAAGGAATTCATTTTTATTCTTTTTACTTATTAAGAGATGATATCTATCAAATTCCAAGCTTA
AAACCAGATTTAGTTATCACTCATAGCCGATTAATTCCTTTTGTTAAGAATGATCTGGTCAAAGGTGTTACTGTT
GCTGAATTTTCTTTTGATAACCCTGACTACTCTATTGCTTCAATTCAAAACTTGATATATCAGCTCAAAGATAAA
AAATATCAAGATTTTCTAAACGAGCAATTACAA

SEQ ID NO: 89

MPYVKKKKDSFLVETYLEQSIRDKSELVLLLFKSPTIIFSHVAKQTGLTAVQLKYYCKELDDFFGNNLDITIKKG
KIICCFVKPVKEFYLHQLYDTSTILKLLVFFIKNGTSSQPLIKFSKKYFLSSSSAYRLRESLIKLLREFGLRVSK
NTIVGEEYRIRYLIAMLYSKFGIVIYPLDHLDNQIIYRFLSQSATNLRTSPWLEEPFSFYNMLLALSWKRHQFAV
SIPQTRIFRQLKKLFIYDCLTRSSRQVIENAFSLMFSQGDLDYLFLIYITTNNSFASLQWTPQHIETCCHIFEKN
DTFRLLLEPILKRLPQLNHSKQDLIKALMYFSKSFLFNLQHFVIEIPSFSLPTYTGNSNLYKALKNIVNQWLAQL
PGKRHLNEKHLQLFCSHIEQILKNKQPALTVVLISSNFINAKLLTDTIPRYFSDKGIHFYSFYLLRDDIYQIPSL
KPDLVITHSRLIPFVKNDLVKGVTVAEFSFDNPDYSIASIQNLIYQLKDKKYQDFLNEQLQ

Orf 78 is thought to be a collagen binding protein (Cbp). An example of the nucleotide sequence encoding Cbp (SEQ ID NO: 90) and a Cbp amino acid sequence (SEQ ID NO: 91) are set forth below.

SEQ ID NO: 90

TTGCAAAAGAGGGATAAAACCAATTATGGAAGCGCTAACAACAAACGACGACAAACGACGATCGGATTACTGAAA
GTATTTTTGACGTTTGTAGCTCTGATAGGAATAGTAGGGTTTTCTATCAGAGCGTTCGGAGCTGAAGAAAAATCT
ACTGAAACTAAAAAAACGTCAGTCATTATTAGAAAATATGCTGAAGGTGACTACTCTAAACTTCTAGAGGGAGCA
ACTTTGCGTTTAACAGGGGAAGATATCCCAGATTTTCAAGAAAAAGTCTTCCAAAGTAATGGAACAGGAGAAAAG
ATTGAATTATCAAATGGGACTTATACCTTAACAGAAACATCATCTCCAGATGGATATAAAATTACGGAGCCGATT

AAGTTTAAGAGTAGTGAATAAAAAGTANTTATCGTCCAAAAAGATGGTTCTCAAGTGGAAAACCCAAACAAAGAA
CTAGGTTCTCCATATACTATAGAGGCATACAATGATTTTGATGAATTTGGCTTACTGTCAACACAAAATTATGCG
AAATTTTATTATGGAAAAAACTATGATGGCAGTTCACAAATTGTTTATTGCTTCAATGCCAACTTGAAATCTCCA
CCTGACTCGGAAGATCATGGTGCTACAATAAATCCTGACTTTACGACTGGTGATATTAGGTACAGTCATATTGCT
GGTTCAGATTTGATAAAATACGCTAATACAGCTAGGGATGAAGATCCTCAATTATTTTTAAAACACGTAAAAAAA
GTAATTGAAAATGGGTATCATAAAAAAGGTCAAGCTATTCCATATAACGGTCTGACTGAGGCACAGTTTCGTGCG
GCTACTCAACTGGCAATTTATTATTTTACAGATAGTGTTGACTTAACTAAGGATAGATTGAAAGACTTCCATGGA
TTTGGAGATATGAATGATCAAACTTTGGGTGTAGCTAAAAAAATTGTAGAATACGCTTTGAGTGATGAAGATTCA
AAACTAACAAATCTTGATTTCTTCGTACCTAATAATAGCAAATACCAATCTCTTATTGGGACAGAATACCATCCA
GATGATTTGGTTGACGTGATTCGTATGGAAGATAAAAAGCAAGAAGTTATTCCAGTAACTCATAGTTTGACGGTG
CAAAAAACAGTAGTCGGTGAGTTGGGAGATAAGACTAAAGGCTTTCAATTTGAACTTGAGTTGAAAGATAAAACT
GGACAGCCTATTGTTAACACTCTAAAAACTAATAATCAAGATTTAGTAGCTAAAGATGGGAAATATTCATTTAAT
CTAAAGCATGGTGACACCATAAGAATAGAAGGATTACCGACGGGATATTCTTATACCCTGAAAGAGACTGAAGCT
AAGGATTATATAGTAACTGTTGATAACAAAGTTAGTCAAGAAGCTCAATCAGCAAGTGAGAATGTCACAGCAGAC
AAAGAAGTCACTTTTGAAAACCGAAAAGATCTTGTCCCACCAACTGGTTTGACAACAGATGGGGCTATCTATCTT
TGGTTATTACTACTTGTTCCATTTGGGTTATTGGTTTGGCTATTTGGTCGTAAAGGGTTAAAAAATGAC

SEQ ID NO: 91

MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEKSTETKKTSVIIRKYAEGDYSKLLEGA
TLRLTGEDIPDFQEKVFQSNGTGEKIELSNGTYTLTETSSPDGYKITEPIKFRVVNKKVFIVQKDGSQVENPNKE
LGSPYTIEAYNDFDEFGLLSTQNYAKFYYGKNYDGSSQIVYCFNANLKSPPDSEDHGATINPDFTTGDIRYSHIA
GSDLIKYANTARDEDPQLFLKHVKKVIENGYHKKGQAIPYNGLTEAQFRAATQLAIYYFTDSVDLTKDRLKDFHG
FGDMNDQTLGVAKKIVEYALSDEDSKLTNLDFFVPNNSKYQSLIGTEYHPDDLVDVIRMEDKKQEVIPVTHSLTV
QKTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYSYTLKETEA
KDYIVTVDNKVSQEAQSASENVTADKEVTFENRKDL*VPPTG*LTTDGAIYLWLLLLVPFGLLVWLFGRKGLKND

Orf 78 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 184** VPPTG (shown in italics in SEQ ID NO: 91, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Orf 78 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Three E boxes containing conserved glutamic residues have been identified in Orf 78. The E-box motifs are underlined in SEQ ID NO: 91, below. The conserved glutamic acid (E) residues, at amino acid residues 112, 395, and 447, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of Orf 78. Preferred fragments of Orf 78 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 91

MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEKSTETKKTSVIIRKYAEGDYSKLLEGA
TLRLTGEDIPDFQEKVFQSNGTGEKIELSNGTYTLTETSSPDGYKITEPIKFRVVNKKVFIVQKDGSQVENPNKE
LGSPYTIEAYNDFDEFGLLSTQNYAKFYYGKNYDGSSQIVYCFNANLKSPPDSEDHGATINPDFTTGDIRYSHIA
GSDLIKYANTARDEDPQLFLKHVKKVIENGYHKKGQAIPYNGLTEAQFRAATQLAIYYFTDSVDLTKDRLKDFHG
FGDMNDQTLGVAKKIVEYALSDEDSKLTNLDFFVPNNSKYQSLIGTEYHPDDLVDVIRMEDKKQEVIPVTHSLTV
QKTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYSYTLKETEA
KDYIVTVDNKVSQEAQSASENVTADKEVTFENRKDLVPPTGLTTDGAIYLWLLLLVPFGLLVWLFGRKGLKND

Orf 79 is thought to be a LepA signal peptidase I. An example of the nucleotide sequence encoding a LepA signal peptidase I (SEQ ID NO: 92) and a LepA signal peptidase I amino acid sequence (SEQ ID NO: 93) are set forth below.

PCT/US05/27239

SEQ ID NO: 92

ATGACTAATTACCTAAATCGTTTAAATGAGAATTCACTATTTAAAGCTTTCATACGGTTAGTACTTAAGATTTCT
ATTATTGGGTTTCTAGGTTACATTCTATTTCAGTATGTTTTTGGTGTTATGATTATTAACACTAATGATATGAGT
CCTGCTTTAAGTGCAGGTGACGGTGTTTTATATTATCGTTTGACTGATCGCTATCATATTAATGATGTGGTGGTC
TATGAGGTTGATAACACTTTGAAAGTTGGTCGAATTGTCGCTCAAGCTGGCGATGAGGTTAGTTTTACGCAAGAA
GGAGGACTGTTGATTAATGGGCATCCACCAGAAAAAGAGGTCCCTTACCTGACGTATCCTCACTCAAGTGGCCCA
AACTTTCCCTATAAAGTTCCTACGGGTAAGTATTTCATATTGAATGATTATCGTGAAGAACGTTTGGACAGTCGT
TATTATGGGGCGTTACCCGTCAATCAAATAAAAGGGAAAATCTCAACTCTATTAAGAGTGAGAGGAATT

SEQ ID NO: 93

MTNYLNRLNENSLFKAFIRLVLKISIIGFLGYILFQYVFGVMIINTNDMSPALSAGDGVLYYRLTDRYHINDVVV
YEVDNTLKVGRIVAQAGDEVSFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGKYFILNDYREERLDSR
YYGALPVNQIKGKISTLLRVRGI

Orf 80 is thought to to be a fimbrial protein. An example of the nucleotide sequence encoding the fimbrial protein (SEQ ID NO: 94) and a fimbrial protein amino acid sequence (SEQ ID NO: 95) are set forth below.

SEQ ID NO: 94

TTGGAGAGAGAAAAAATGAAAAAAAACAAATTATTACTTGCTACTGCAATCTTAGCAACTGCTTTAGGAACAGCT
TCTTTAAATCAAAACGTAAAAGCTGAGACGGCAGGGGTTGTAACAGGAAAATCACTACAAGTTACAAAGACAATG
ACTTATGATGATGAAGAGGTGTTAATGCCCGAAACCGCCTTTACTTTTACTATAGAGCCTGATATGACTGCAAGT
GGAAAAGAAGGCAGCCTAGATATTAAAAATGGAATTGTAGAAGGCTTAGACAAACAAGTAACAGTAAAATATAAG
AATACAGATAAAACATCTCAAAAAACTAAAATAGCACAATTTGATTTTTCTAAGGTTAAATTTCCAGCTATAGGT
GTTTACCGCTATATGGTTTCAGAGAAAAACGATAAAAAAGACGGAATTACGTACGATGATAAAAAGTGGACTGTA
GATGTTTATGTTGGGAATAAGGCCAATAACGAAGAAGGTTTCGAAGTTCTATATATTGTATCAAAAGAAGGTACT
TCTAGTACTAAAAAACCAATTGAATTTACAAACTCTATTAAAACTACTTCCTTAAAAATTGAAAAACAAATAACT
GGCAATGCAGGAGATCGTAAAAAATCATTCAACTTCACATTAACATTACAACCAAGTGAATATTATAAAACTGGA
TCAGTTGTGAAAATCGAACAGGATGGAAGTAAAAAAGATGTGACGATAGGAACGCCTTACAAATTTACTTTGGGA
CACGGTAAGAGTGTCATGTTATCGAAATTACCAATTGGTATCAATTACTATCTTAGTGAAGACGAAGCGAATAAA
GACGGCTACACTACAACGGCAACATTAAAAGAACAAGGCAAAGAAAAGAGTTCCGATTTCACTTTGAGTACTCAA
AACCAGAAAACAGACGAATCTGCTGACGAAATCGTTGTCACAAATAAGCGTGACACTCAAGTTCCAACTGGTGTT
GTAGGGACCCTTGCTCCATTTGCAGTTCTTAGCATTGTGGCTATTGGTGGAGTTATCTATATTACAAAACGTAAA
AAAGCT

SEQ ID NO: 95

MEREKMKKNKLLLATAILATALGTASLNQNVKAETAGVVTGKSLQVTKTMTYDDEEVLMPETAFTFTIEPDMTAS
GKEGSLDIKNGIVEGLDKQVTVKYKNTDKTSQKTKIAQFDFSKVKFPAIGVYRYMVSEKNDKKDGITYDDKKWTV
DVYVGNKANNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLKIEKQITGNAGDRKKSFNFTLTLQPSEYYKTG
SVVKIEQDGSKKDVTIGTPYKFTLGHGKSVMLSKLPIGINYYLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQ
NQKTDESADEIVVTNKRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

Orf 82 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 140** QVPTG (shown in italics in SEQ ID NO: 95, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Orf 82 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

An E box containing a conserved glutamic residue has been identified in Orf 80. The E-box motif is underlined in SEQ ID NO: 95, below. The conserved glutamic acid (E), at amino acid residue 270, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is

thought to be important for the formation of oligomeric pilus-like structures of Orf 80. Preferred fragments of Orf 80 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 95

MEREKMKKNKLLLATAILATALGTASLNQNVKAETAGVVTGKSLQVTKTMTYDDEEVLMPETAFTFTIEPDMTAS
GKEGSLDIKNGIVEGLDKQVTVKYKNTDKTSQKTKIAQFDFSKVKFPAIGVYRYMVSEKNDKKDGITYDDKKWTV
DVYVGNKANNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLKIEKQITGNAGDRKKSFNFTLTLQPSEYYKTG
SVVKIEQDGSKKDVTIGTPYKFTLGHGKSVMLSKLPIGINYYLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQ
NQKTDESADEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

Orf 81 is thought to to be a SrtC2 type sortase. An example of the nucleotide sequence encoding the SrtC2 sortase (SEQ ID NO: 96) and a SrtC2 sortase amino acid sequence (SEQ ID NO: 97) are set forth below.

SEQ ID NO: 96

GTGATTAGTCAAAGAATGATGATGACAATTGTACAGGTTATCAATAAAGCCATTGATACTCTCATTCTTATCTTT
TGTTTAGTCGTACTATTTTTAGCTGGTTTTGGTTTGTGGGATTCTTATCATCTCTATCAACAAGCAGACGCTTCT
AATTTCAAAAAATTTAAAACAGCTCAACAACAGCCTAAATTTGAAGACTTGTTAGCTTTGAATGAGGATGTCATT
GGTTGGTTAAATATCCCAGGGACTCATATTGATTATCCTCTAGTTCAGGGAAAAACGAATTTAGAGTATATTAAT
AAAGCAGTTGATGGCAGTGTTGCCATGTCTGGTAGTTTATTTTTAGATACACGGAATCATAATGATTTTACGGAC
GATTACTCTCTGATTTATGGCCATCATATGGCAGGTAATGCCATGTTTGGCGAAATTCCAAAATTTTTAAAAAAG
GATTTTTTCAACAAACATAATAAAGCTATCATTGAAACAAAAGAGAGAAAAAAACTAACCGTCACTATTTTTGCT
TGTCTCAAGACAGATGCCTTTGACCAGTTAGTTTTTAATCCTAATGCTATTACCAATCAAGACCAACAAAAGCAG
CTCGTTGATTATATCAGTAAAAGATCAAAACAATTTAAACCTGTTAAATTGAAGCATCATACAAAGTTCGTTGCT
TTTTCAACGTGTGAAAATTTTTCTACTGACAATCGTGTTATCGTTGTCGGTACTATTCAAGAA

SEQ ID NO: 97

MISQRMMMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVI
GWLNIPGTHIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIPKFLKK
DFFNKHNKAIIETKERKKLTVTIFACLKTDAFDQLVFNPNAITNQDQQKQLVDYISKRSKQFKPVKLKHHTKFVA
FSTCENFSTDNRVIVVGTIQE

Orf 82 is referred to as a hypothetical protein. It contains a sortase substrate motif LPXAG shown in italics in SEQ ID NO: 99. An example of the nucleotide sequence encoding the hypothetical protein (SEQ ID NO: 98) and a hypothetical protein amino acid sequence (SEQ ID NO: 99) are set forth below.

SEQ ID NO: 98

TTGCTTTTTCAACGTGTGAAAATTTTTCTACTGACAATCGTGTTATCGTTGTCGGTACTATTCAAGAATAACGAA
AGGAGGAGACTTTTGAGAAAATATTGGAAAATGTTATTTTCTGTCGTAATGATATTAACCATGCTGGCCTTTAAT
CAGACTGTTTTAGCAAAAGACAGCACTGTTCAAACTAGCATTAGTGTCGAAAATGTCTTAGAGAGAGCAGGCGAT
AGTACCCCATTTTCGGTTGCATTAGAATCAATTGATGCGATGAAAACAATAGACGAAATAACAATTGCTGGTTCT
GGAAAAGCAAGCTTTTCCCCTCTGACCTTCACAACAGTTGGGCAATATACTTATCGTGTTTATCAGAAGCCTTCA
CAAAATAAAGATTATCAAGCAGATACTACTGTATTTGACGTTCTTGTCTATGTGACCTATGATGAAGATGGGACT
CTAGTCGCAAAAGTTATTTCTCGAAGGGCTGGAGACGAAGAAAAATCAGCGATTACTTTTAAGCCCAAACGGTTA
GTAAAACCAATACCGCCTAGACAACCTAACATCCCTAAAACCCCATTACCATTAGCTGGTGAAGTAAAAAGTTTA
TTGGGTATCTTAAGTATCGTATTACTGGGGTTACTAGTTCTTCTTTATGTTAAAAAACTGAAGAGTAGGCTA

SEQ ID NO: 99

MLFQRVKIFLLTIVLSLSVLFKNNERRRLLRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGD
STPFSVALESIDAMKTIDEITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGT
LVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTP*LPLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSRL

P1C Orf 82 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 185** LPLAG (shown in italics in SEQ ID NO: 99, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Orf 82 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in Orf 82. The pilin motif sequence is underlined in SEQ ID NO: 99, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 173 and 188. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of Orf 82 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 99

MLFQRVKIFLLTIVLSLSVLFKNNERRRLLRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGD
STPFSVALESIDAMKTIDEITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGT
LVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

An E box containing a conserved glutamic residue has been identified in Orf 82. The E-box motif is underlined in SEQ ID NO: 99, below. The conserved glutamic acid (E), at amino acid residue 163, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of Orf 82. Preferred fragments of Orf 82 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 99

MLFQRVKIFLLTIVLSLSVLFKNNERRRLLRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGD
STPFSVALESIDAMKTIDEITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGT
LVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

Orf 83 is thought to to be a multiple sugar metabolism regulator protein. An example of a nucleotide sequence encoding the sugar metabolism regulator protein (SEQ ID NO: 100) and a sugar metabolism regulator protein amino acid sequence (SEQ ID NO: 101) are set forth below.

SEQ ID NO: 100

ATGATACAACTAAGGATGGGGGCAATCTATCAAATGGTTATATTCGATTTAAAACATGTGCAAACATTACACAGC
TTGTCTCAATTACCTATTTCAGTGATGTCACAAGATAAGGCACTTATTCAAGTATATGGTAATGACGACTATTTA
TTATGTTACTATCAATTTTTAAAGCATCTAGCTATTCCTCAAGCTGCACAAGATGTTATTTTTTATGAGGGTTTA
TTTGAAGAGTCCTTTATGATTTTTCCTCTTTGTCACTACATTATTGCCATTGGACCTTTCTATCCTTATTCACTT
AATAAAGACTATCAGGAACAATTAGCTAATAATTTTTTAAAACATTCTTCTCATCGTAGCAAAGAAGAGCTCTTG
TCCTATATGGCACTTGTCCCACATTTTCCAATTAATAATGTGCGGAACCTTTTGATAGCTATTGACGCTTTTTTT
GACACACAATTTGAGACGACTTGCCAACAAACGATTCATCAATTGTTGCAGCATTCAAAACAGATGACTGCTGAT
CCTGATATCATTCATCGCCTTAAGCATATTAGCAAAGCATCTAGCCAATTACCGCCTGTTTTAGAGCACCTAAAT
CATATTATGGATCTGGTAAAGCTAGGCAATCCACAATTGCTCAAGCAAGAAATCAATCGCATCCCCTTATCAAGT
ATCACCTCATCTTCTATTTCTGCTCTAAGGGCGGAAAAGAACCTCACTGTTATCTATTTAACTAGGTTACTGGAA
TTCAGTTTTGTAGAAAATACTGACGTAGCAAAGCATTATAGCCTTGTCAAATACTACATGGCCTTAAATGAAGAA
GCGAGTGACTTGCTCAAAGTTTTGAGAATTCGCTGTGCAGCTATCATCCATTTTTCCGAATCATTAACCAATAAA
AGTATTTCTGATAAACGTCAAATGTACAATAGTGTGCTTCATTATGTCGATAGTCACCTGTATTCCAAATTAAAG

GTATCTGATATCGCTAAGCGCCTATATGTTTCCGAATCTCACTTACGTTCAGTCTTTAAAAAATACTCAAATGTT
TCCTTACAACATTATATTCTAAGTACAAAAATCAAAGAAGCTCAACTACTCTTAAAACGAGGAATTCCTGTTGGA
GAAGTGGCTAAAAGCTTATATTTTTATGACACTACCCATTTTCATAAAATCTTTAAAAAATACACGGGTATTTCT
TCAAAAGACTATCTTGCTAAATACCGAGATAATATT

SEQ ID NO: 101

MIQLRMGAIYQMVIFDLKHVQTLHSLSQLPISVMSQDKALIQVYGNDDYLLCYYQFLKHLAIPQAAQDVIFYEGL
FEESFMIFPLCHYIIAIGPFYPYSLNKDYQEQLANNFLKHSSHRSKEELLSYMALVPHFPINNVRNLLIAIDAFF
DTQFETTCQQTIHQLLQHSKQMTADPDIIHRLKHISKASSQLPPVLEHLNHIMDLVKLGNPQLLKQEINRIPLSS
ITSSSISALRAEKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEASDLLKVLRIRCAAIIHFSESLTNK
SISDKRQMYNSVLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSLQHYILSTKIKEAQLLLKRGIPVG
EVAKSLYFYDTTHFHKIFKKYTGISSKDYLAKYRDNI

Orf 84 is thought to to be a F2-like fibronectin-binding protein. An example of a nucleotide sequence encoding the F2-like fibronectin-binding protein (SEQ ID NO: 102) and a F2-like fibronectin-binding protein amino acid sequence (SEQ ID NO: 103) are set forth below.

SEQ ID NO: 102

ATGACACAAAAAAATAGCTATAAGTTAAGCTTCCTGTTATCCCTAACAGGATTTATTTTAGGTTTATTATTGGTT
TTTATAGGATTGTCCGGAGTATCAGTAGGACATGCGGAAACAAGAAATGGAGCAAACAAACAAGGAGCTTTTGAA
ATCAAGAAAAATAAAAGTCAAGAAGAATATAATTATGAAGTTTATGATAACAGAAACATACTTCAGGATGGGGAA
CATAAACTTGAAATAAAAAGAGTTGATGGGACAGGTAAAACTTATCAAGGTTTTTGCTTTCAGTTAACGAAAAAT
TTTCCCACTGCTCAAGGTGTAAGTAAAAAGCTGTATAAAAAATTGAGTAGTAGTGATGAAGAAACACTAAAGCAA
TATGCCTCTAAGTATACAAGTAATAGGAGAGGAGATACTAGTGGTAATCTTAAAAAGCAAATTGCTAAGGTTCTG
ACAGAAGGTTACCCAACTAACAAAAGTGATTGGTTAAATGGATTGACTGAAAACGAAAAAATAGAAGTAACCCAG
GATGCAATTTGGTATTTTACAGAAACGACAGTTCCGGCTGATAGAAGTTATACGAATCGCAACGTAAATAGTCAA
AAAATGAAAGAAGTGTATCAAAAGCTAATTGATACAACAGATATAGATAAATATGAAGATGTACAATTTGATTTA
TTTGTGCCACAAGATACAAACTTACAGGCAGTAATTAGTGTAGAGCCTGTTATCGAAAGCCTTCCTTGGACATCG
TTGAAGCCAATAGCCCAGAAGGATATCACTGCCAAAAAAATCTGGGTAGATGCACCTAAAGAAAAACCAATTATT
TATTTTAAGCTATATAGACAGCTGCCTGGAGAAAAGGAAGTAGCAGTGGATGACGCTGAGCTAAAACAGATAAAT
AGTGAAGGTCAACAAGAAATATCAGTAACTTGGACAAATCAACTTGTTACAGATGAAAAAGGAATGGCTTACATT
TATTCTGTAAAAGAAGTAGATAAAAATGGCGAGTTACTTGAGCCAAAAGATTATATCAAGAAGGAAGATGGACTT
ACAGTTACTAATACTTATGTAAAGCCAACTAGTGGGCACTATGATATAGAAGTGACATTTGGAAATGGACATATT
GATATTACAGAAGATACTACACCAGATATTGTTTCAGGTGAAAACCAAATGAAGCAAATAGAGGGAGAAGATAGT
AAGCCTATTGATGAAGTAACGGAAAATAATTTAATTGAATTTGGTAAAAACACGATGCCAGGTGAAGAAGATGGC
ACAAATTCTAATAAGTATGAAGAAGTCGAAGACTCACGCCCAGTTGATACCTTGTCAGGTTTATCAAGTGAGCAA
GGTCAGTCCGGTGATATGACAATTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGATATTGACGGC
AAAGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACATGGATTTCAGATGGA
CAAGTGAAAGATTTCTACCTGATGCCAGGAAAATATACATTTGTCGAAACCGCAGCACCAGACGGTTATGAGATA
GCAACTGCTATTACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGACGCT
CATATTGTCATGGTTGATGCTTACAAGCCAACTAAGGGTTCAGGTCAGGTTATTGATATTGAAGAAAAGCTTCCA
GACGAGCAGGGCCATTCTGGCTCAACTACTGAAATAGAAGATAGCAAGTCTTCAGACGTTATCATTGGTGGTCAG
GGGCAGATTGTCGAGACAACAGAGGATACCCAAACTGGCATGCACGGGGATTCTGGTTGTAAAACGGAAGTCGAA
GATACTAAACTAGTACAATCCTTCCACTTTGATAACAAGGAATCAGAAAGTAACTCTGAGATTCCTAAAAAAGAT
AAGCCAAAGAGTAATACTAGTTTACCAGCAACTGGTGAGAAGCAACATAATATGTTCTTTTGGATGGTTACTTCT
TGCTCACTTATTAGTAGTGTTTTTGTAATATCACTAAAAACTAAAAAACGCCTATCATCATGT

SEQ ID NO: 103

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTS*LPATG*EKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

Orf 84 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 181** LPATG (shown in italics in SEQ ID NO: 103, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant Orf 84 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in Orf 84. The pilin motif sequence is underlined in SEQ ID NO: 103, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 270. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of Orf 84 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 103

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTSLPATGEKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

An E box containing a conserved glutamic residue has been identified in Orf 84. The E-box motif is underlined in SEQ ID NO: 103, below. The conserved glutamic acid (E), at amino acid residue 516, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of Orf 84. Preferred fragments of Orf 84 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 103

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTSLPATGEKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

Examples of GAS AI-3 sequences from M18 strain isolate MGAS8232 are set forth below.

SpyM18_0125 is a negative transcriptional regulator (Nra). An example of SpyM18_0125 is set forth in SEQ ID NO: 72.

SEQ ID NO: 72

MPYVKKKKDSFEVETYLEQSIRDKSELVLLFKSPTIIFSHVAKQTGLTAVQLKYYCKELDDFFGNNLDITIKKG
KIICCFVKPVKEFYLHQLYDTSTILKLLVFFIKNGTTSQPLIKFSKKYFLSSSSAYRLRESLIKLLREFGLRVSK
NTIVGEEYRIRYLIAMLYSKFGIVIYPLDHLDNQIIYRFLSQSATNLRTSPWLEEPFSFYNMLLALS

SpyM18_0126 is thought to be a collagen binding protein (CBP). An example of SpyM18_0126 is set forth in SEQ ID NO: 73.

SEQ ID NO: 73
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSTETKKTSVIIRKYAEGDYSKLLEGA
TLKLAQIEGSGFQEQSFESSTSGQKLQLSDGTYILTETKSPQGYEIAEPITFKVTAGKVFIKGKDGQFVENQNKE
VAEPYSVTAYNDFDDSGFINPKTFTPYGKFYYAKNANGTSQVVYCFNVDLHSPPDSLDKGETIDPDFNEGKEIKY
THILGADLFSYANNPRASTNDELLSQVKKVLEKGYRDDSTTYANLTSVEFRAATQLAIYYFTDSVDLDNLADYHG
FGALTTEALNATKEIVAYAEDRANLPNISNLDFYVPNSNKYQSLIGTQYHPESLVDIIRMEDKQAPIIPITHKLT
ISKTVTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETG
ASDYEVSVNGKNAPDGKATKASVKEDETITFENRKDL*VPPTG*LTTDGAIYLWLLLLVLLGLWVWLIGRKGLKND

SpyM18_0126 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 184** VPPTG (shown in italics in SEQ ID NO: 73, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM18_0126 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM18_0126. The pilin motif sequence is underlined in SEQ ID NO: 73, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 172 and 179. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM18_0126 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 73
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSTETKKTSVIIRKYAEGDYSKLLEGA
TLKLAQIEGSGFQEQSFESSTSGQKLQLSDGTYILTETKSPQGYEIAEPITFKVTAGKVFIKGKDGQFVENQNKE
VAEPYSVTAYNDFDDSGFINPKTFTPYGKFYYAKNANGTSQVVYCFNVDLHSPPDSLDKGETIDPDFNEGKEIKY
THILGADLFSYANNPRASTNDELLSQVKKVLEKGYRDDSTTYANLTSVEFRAATQLAIYYFTDSVDLDNLADYHG
FGALTTEALNATKEIVAYAEDRANLPNISNLDFYVPNSNKYQSLIGTQYHPESLVDIIRMEDKQAPIIPITHKLT
ISKTVTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETG
ASDYEVSVNGKNAPDGKATKASVKEDETITFENRKDLVPPTGLTTDGAIYLWLLLLVLLGLWVWLIGRKGLKND

Three E boxes containing conserved glutamic residues have been identified in SpyM18_0126. The E-box motifs are underlined in SEQ ID NO: 73, below. The conserved glutamic acid (E) residues, at amino acid residues 112, 257, and 415, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyM18_0126. Preferred fragments of SpyM18_0126 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 73
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSTETKKTSVIIRKYAEGDYSKLLEGA
TLKLAQIEGSGFQEQSFESSTSGQKLQLSDGTYILTETKSPQGYEIAEPITFKVTAGKVFIKGKDGQFVENQNKE

VAEPTSVTAYNDFDDGGETNPRTFTPYGKFLAKNANGTSQVVYCFNVDLHSPPDSLDKGETIDPDFNEGKEIKY
THILGADLFSYANNPRASTNDELLSQVKKVLEKGYRDDSTTYANLTSVEFRAATQLAIYYFTDSVDLDNLADYHG
FGALTTEALNATKEIVAYAEDRANLPNISNLDFYVPNSNKYQSLIGTQYHPESLVDIIRMEDKQAPIIPITHKLT
ISKTVTGTIADKKKEFNFEIHLKSSDGQAISGTYPTNSGELTVTDGKATFTLKDGESLIVEGLPSGYSYEITETG
ASDYEVSVNGKNAPDGKATKASVKEDETITFENRKDLVPPTGLTTDGAIYLWLLLLVLLGLWVWLIGRKGLKND

SpyM18_0127 is a LepA protein. An example of SpyM18_0127 is shown in SEQ ID NO: 74.

SEQ ID NO: 74

MTNYLNRLNENPLFKAFIRLVLKISIIGFLGYILFQYIFGVMIINTNVMSPALSAGDGILYYRLTDRYHINDVVV
YEVDNTLKVGRIVAQAGDEVSFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILNDYREERLDSR
YYGALPINQIKGKISTLLRVRGI

SpyM18_0128 is thought to be a fimbrial protein. An example of SypM18_0128 is shown in SEQ ID NO: 75.

SEQ ID NO: 75

MKKNKLLLATAILATALGTASLNQNVKAETAGVIDGSTLVVKKTFPSYTDDKVLMPKADYTFKVEADDNAKGKTK
DGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFDFANVKFPGVGVYRYTVSEVNGNKAGIAYDSQQWTVDVY
VVNREDGGFEAKYIVSTEGGQSDKKPVLFKNFFDTTSLKVTKKVTGNTGEHQRSFSFTLLLTPNECFEKGQVVNI
LQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKVTEEDVTKDGYKTSATLKDGDVTDGYNLGDSKTTDKST
DEIVVTNKRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyM18_0128 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 140** QVPTG (shown in italics in SEQ ID NO: 75, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM18_0128 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM18_0128. The pilin motif sequence is underlined in SEQ ID NO: 75, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 57. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM18_0128 include the conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 75

MKKNKLLLATAILATALGTASLNQNVKAETAGVIDGSTLVVKKTFPSYTDDKVLMPKADYTFKVEADDNAKGKTK
DGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFDFANVKFPGVGVYRYTVSEVNGNKAGIAYDSQQWTVDVY
VVNREDGGFEAKYIVSTEGGQSDKKPVLFKNFFDTTSLKVTKKVTGNTGEHQRSFSFTLLLTPNECFEKGQVVNI
LQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKVTEEDVTKDGYKTSATLKDGDVTDGYNLGDSKTTDKST
DEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

An E box containing a conserved glutamic residue has been identified in SpyM18_0128. The E-box motif is underlined in SEQ ID NO: 75, below. The conserved glutamic acid (E), at amino acid residue 266, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyM18_0128.

Preferred fragments of SpyM18_0128 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 75

MKKNKLLLATAILATALGTASLNQNVKAETAGVIDGSTLVVKKTFPSYTDDKVLMPKADYTFKVEADDNAKGKTK
DGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFDFANVKFPGVGVYRYTVSEVNGNKAGIAYDSQQWTVDVY
VVNREDGGFEAKYIVSTEGGQSDKKPVLFKNFFDTTSLKVTKKVTGNTGEHQRSFSFTLLLTPNECFEKGQVVNI
LQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKVTEEDVTKDGYKTSATLKDGDVTDGYNLGDSKTTDKST
DEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyM18_0129 is a SrtC2 type sortase. An example of SpyM18_0129 is shown in SEQ ID NO: 76

SEQ ID NO: 76

MISQRMMMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVI
GWLNIPGTHMDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIPKFLKK
DFFNKHNKAIIETKERKKLTVTIFACLKTDAFDQLVFNPNAITNQDQQRQLVDYISKRSKQFKPVKLKHHTKFVA
FSTCENFSTDNRVIVVGTIQE

SpyM18_0130 is referred to as a hypothetical protein. An example of SpyM18_0130 is shown in SEQ ID NO: 77.

SEQ ID NO: 77

MRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTSFSVALESIDAMKTIDEITIAGSGKAS
FSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPI
PPRQPDIPKTP*LPLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyM18_0130 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 185** LPLAG (shown in italics in SEQ ID NO: 77, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM18_0130 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM18_0130. The pilin motif sequence is underlined in SEQ ID NO: 77, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 144, 159, and 169. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM18_0130 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 77

MRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTSFSVALESIDAMKTIDEITIAGSGKAS
FSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPI
PPRQPDIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

An E box containing a conserved glutamic residue has been identified in SpyM18_0130. The E-box motif is underlined in SEQ ID NO: 77, below. The conserved glutamic acid (E), at amino acid residue 134, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is

thought to be important for the formation of oligomeric pilus-like structures of SpyM18_0130. Preferred fragments of SpyM18_0130 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 77

MRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTSFSVALESIDAMKTIDEITIAGSGKAS
FSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPI
PPRQPDIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyM18_0131 is referred to as a putative multiple sugar metabolism regulator. An example of SpyM18_0131 is set forth in SEQ ID NO: 78.

SEQ ID NO: 78

MAIFDLKHVQTLHSLSQLPISVMSQDKALIQVYGNDDYLLCYYQFLKHLAIPQAAQDVIFYEGLFEESFMIFPLC
HYIIAIGPFYPYSLNKDYQEQLANNCLKHSSHRSKEELLSYMALVPHFPINNVRNLLIAIDAFFDTQFETTCQQT
IHQLLQHSKQMTADPDIIHRLKHISKASSQLPPVLEHLNHIMDLVKLGNPQLLKQEINRIPLSSITSSSISALRA
EKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEASDLLKVLRIRCAAIIHFSESLTNKSISDKRQMYNS
VLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSLQHYILSTKIKEAQLLLKRGIPVGEVAKSLYFYDT
THFHKIFKKYTGISSKDYLAKYRDNI

SpyM18_0132 is a F2 like fibronectic-binding protein. An example of SpyM18_0132 is set forth in SEQ ID NO: 79.

SEQ ID NO: 79

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTS*LPATG*EKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

SpyM18_0132 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 180** LPATG (shown in italics in SEQ ID NO: 79, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyM18_0132 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyM18_0132. The pilin motif sequence is underlined in SEQ ID NO: 79, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 270. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyM18_0132 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 79

MTQKNSYKESFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTSLPATGEKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

An E box containing a conserved glutamic residue has been identified in SpyM18_0132. The E-box motif is underlined in SEQ ID NO: 79, below. The conserved glutamic acid (E), at amino acid residue 516, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyM18_0132. Preferred fragments of SpyM18_0132 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 79

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGAFEIKKNKSQEEYNYEVYDNRNILQDGE
HKLEIKRVDGTGKTYQGFCFQLTKNFPTAQGVSKKLYKKLSSSDEETLKQYASKYTSNRRGDTSGNLKKQIAKVL
TEGYPTNKSDWLNGLTENEKIEVTQDAIWYFTETTVPADRSYTNRNVNSQKMKEVYQKLIDTTDIDKYEDVQFDL
FVPQDTNLQAVISVEPVIESLPWTSLKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEVAVDDAELKQIN
SEGQQEISVTWTNQLVTDEKGMAYIYSVKEVDKNGELLEPKDYIKKEDGLTVTNTYVKPTSGHYDIEVTFGNGHI
DITEDTTPDIVSGENQMKQIEGEDSKPIDEVTENNLIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLSGLSSEQ
GQSGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEI
ATAITFTVNEQGQVTVNGKATKGDAHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQ
GQIVETTEDTQTGMHGDSGCKTEVEDTKLVQSFHFDNKESESNSEIPKKDKPKSNTSLPATGEKQHNMFFWMVTS
CSLISSVFVISLKTKKRLSSC

Examples of GAS AI-3 sequences from M49 strain isolate 591 are set forth below.

SpyoM01000156 is a negative transcriptional regulator (Nra). An example of SpyoM01000156 is set forth in SEQ ID NO: 243.

SEQ ID NO: 243

MPYVKKKKDSFLVETYLEQSIRDKSELVLLLFKSPTIIFSHVAKQTGLTAVQLKYYCKELDDFFGNNLDI
TIKKGKIICCFVKPVKEFYLHQLYDTSTILKLLVFFIKNGTSSQPLIKFSKKYFLSSSSAYRLRESLIKL
LREFGLRVSKNTIVGEEYRIRYLIAMLYSKFGIVIYPLDHLDNQIIYRFLSQSATNLRTSPWLEEPFSFY
NMLLALSWKRHQFAVSIPQTRIFRQLKKLFIYDCLTRSSRQVIENAFSLTFSQGDLDYLFLIYITTNNSF
ASLQWTPQHIETCCHIFEKNDTFRLLLEPILKRLPQLNHSKQDLIKALMYFSKSFLFNLQHFVIEIPSFS
LPTYTGNSNLYKALKNIVNQWLAQLPGKRHLNEKHLQLFCSHIEQILKNKQPALTVVLISSNFINAKLLT
DTIPRYFSDKGIHFYSFYLLRDDIYQIPSLKPDLVITHSRLIPFVKNDLVKGVTVAEFSFDNPDYSIASI
QNLIYQLKDKKYQDFLNEQLQ

SpyoM01000155 is thought to be a collagen binding protein (CPA). An example of SpyoM01000155 is set forth in SEQ ID NO: 244.

SEQ ID NO: 244

MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNRQSSIQDYPWYGYDSYP
KGYPDYSPLKTYHNLKVNLEGSKDYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDG
QLQQNILRILYNGYPNNRNGIMKGIDPLNAILVTQNAIWYYTDSAQINPDESFKTEARSNGINDQQLGLM
RKALKELIDPNLGSKYSNKTPSGYRLNVFESHDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKY
AEGDYSKLLEGATLKLSQIEGSGFQEKDFQSNSLGETVELPNGTYTLTETSSPDGYKIAEPIKFRVENKK
VFIVQKDGSQVENPNKEVAEPYSVEAYNDFMDEEVLSGFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPD
SYDSGETINPDTSTMKEVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYKKKGDSYNGLTETQF
RAATQLAIYYFTDSADLKTLKTYNNGKGYHGFESMDEKTLAVTKELITYAQNGSAPQLTNLDFFVPNNSK
YQSLIGTEYHPDDLVDVIRMEDKKQEVIPVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLK

TNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYSYTLKETEAKDYIVTVDNKVSQEAQSVGKDITEDKKVT
FENRKDL*VPPTG*LTTDGAIYLWLLLLVPLGLLVWLFGRKGLKND

SpyoM01000155 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 184** VPPTG (shown in italics in SEQ ID NO: 244, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyoM01000155 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, discussed above, containing conserved lysine (K) residues have also been identified in SpyoM01000155. The pilin motif sequence is underlined in SEQ ID NO: 244, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 71 and 261. The pilin sequences, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyoM01000155 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 244
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNRQSSIQDYPWYGYDSYP
KGYPDYSPLKTYHNLKVNLEGSKDYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDG
QLQQNILRILYNGYPNNRNGIMKGIDPLNAILVTQNAIWYYTDSAQINPDESFKTEARSNGINDQQLGLM
RKALKELIDPNLGSKYSNKTPSGYRLNVFESHDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKY
AEGDYSKLLEGATLKLSQIEGSGFQEKDFQSNSLGETVELPNGTYTLTETSSPDGYKIAEPIKFRVENKK
VFIVQKDGSQVENPNKEVAEPYSVEAYNDFMDEEVLSGFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPD
SYDSGETINPDTSTMKEVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYKKKGDSYNGLTETQF
RAATQLAIYYFTDSADLKTLKTYNNGKGYHGFESMDEKTLAVTKELITYAQNGSAPQLTNLDFFVPNNSK
YQSLIGTEYHPDDLVDVIRMEDKKQEVIPVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLK
TNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYSYTLKETEAKDYIVTVDNKVSQEAQSVGKDITEDKKVT
FENRKDLVPPTGLTTDGAIYLWLLLLVPLGLLVWLFGRKGLKND

Two E boxes containing conserved glutamic residues have been identified in SpyoM01000155. The E-box motifs are underlined in SEQ ID NO: 244, below. The conserved glutamic acid (E) residues, at amino acid residues 329 and 668, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyoM01000155. Preferred fragments of SpyoM01000155 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 244
MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIGIVGFSIRAFGAEEQSVPNRQSSIQDYPWYGYDSYP
KGYPDYSPLKTYHNLKVNLEGSKDYQAYCFNLTKHFPSKSDSVRSQWYKKLEGTNENFIKLADKPRIEDG
QLQQNILRILYNGYPNNRNGIMKGIDPLNAILVTQNAIWYYTDSAQINPDESFKTEARSNGINDQQLGLM
RKALKELIDPNLGSKYSNKTPSGYRLNVFESHDKTFQNLLSAEYVPDTPPKPGEEPPAKTEKTSVIIRKY
AEGDYSKLLEGATLKLSQIEGSGFQEKDFQSNSLGETVELPNGTYTLTETSSPDGYKIAEPIKFRVENKK
VFIVQKDGSQVENPNKEVAEPYSVEAYNDFMDEEVLSGFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPD
SYDSGETINPDTSTMKEVKYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYKKKGDSYNGLTETQF

RAATQLATYYFTDSADLKTLKTNNGKGYHGTESMDEKTLAVTKELITYAQNGSAPQLTNLDFFVPNNSK
YQSLIGTEYHPDDLVDVIRMEDKKQEVIPVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLK
TNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYSYTLKETEAKDYIVTVDNKVSQEAQSVGKDITEDKKVT
FENRKDLVPPTGLTTDGAIYLWLLLLVPLGLLVWLFGRKGLKND

SpyoM01000154 is a LepA protein. An example of SpyoM01000154 is shown in SEQ ID NO: 245.

SEQ ID NO: 245

MTNYLNRLNENSLFKAFIRLVLKISIIGFLGYILFQYVFGVMIINTNDMSPALSAGDGVLYYRLADRSHI
NDVVVYEVDNTLKVGRIAAQAGDEVNFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILN
DYREERLDSRYYGALPINQIKGKISTLLRVRGI

SpyoM01000153 is thought to be a fimbrial protein. An example of SpyoM01000153 is shown in SEQ ID NO: 246.

SEQ ID NO: 246

MKKNKLLLATAILATALGMASMSQNIKAETAGVIDGSTLVVKKTFPSYTDDNVLMPKADYSFKVEADDNA
KGKTKDGLDIKPGVIDGLENTKTIRYSNSDKITAKEKSVNFEFANVKFPGVGVYRYTVAEVNGNKAGITY
DSQQWTVDVYVVNKEGGGFEVKYIVSTEVGQSEKKPVLFKNSFDTTSLKIEKQVTGNTGEHQRLFSFTLL
LTPNECFEKGQVVNILQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKLTEEDVTKDGYKTSATLK
DGEQSSTYELGKDHKTDKSADEIVVTNKRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyoM01000153 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 140** QVPTG (shown in italics in SEQ ID NO: 246, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyoM01000153 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyoM01000153. The pilin motif sequence is underlined in SEQ ID NO: 246, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 57. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyoM01000153 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 246

MKKNKLLLATAILATALGMASMSQNIKAETAGVIDGSTLVVKKTFPSYTDDNVLMPKADYSFKVEADDNA
KGKTKDGLDIKPGVIDGLENTKTIRYSNSDKITAKEKSVNFEFANVKFPGVGVYRYTVAEVNGNKAGITY
DSQQWTVDVYVVNKEGGGFEVKYIVSTEVGQSEKKPVLFKNSFDTTSLKIEKQVTGNTGEHQRLFSFTLL
LTPNECFEKGQVVNILQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKLTEEDVTKDGYKTSATLK
DGEQSSTYELGKDHKTDKSADEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

An E box containing a conserved glutamic residue has been identified in SpyoM01000153. The E-box motif is underlined in SEQ ID NO: 246, below. The conserved glutamic acid (E), at amino acid residue 265, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of SpyoM01000153. Preferred fragments of SpyoM01000153 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 246

MKKNKLLLATAILATALGMASMSQNIKAETAGVIDGSTLVVKKTFPSYTDDNVLMPKADYSFKVEADDNA
KGKTKDGLDIKPGVIDGLENTKTIRYSNSDKITAKEKSVNFEFANVKFPGVGVYRYTVAEVNGNKAGITY
DSQQWTVDVYVVNKEGGGFEVKYIVSTEVGQSEKKPVLFKNSFDTTSLKIEKQVTGNTGEHQRLFSFTLL
LTPNECFEKGQVVNILQGGETKKVVIGEEYSFTLKDKESVTLSQLPVGIEYKLTEEDVTKDGYKTSATLK
DGEQSSTYELGKDHKTDKSADEIVVTNKRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

SpyoM01000152 is a SrtC2 type sortase. An example of SpyoM01000152 is shown in SEQ ID NO: 247

SEQ ID NO: 247

MMMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVI
GWLNIPGTHIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIP
KFLKKNFFNKHNKAIIETKERKKLTVTIFACLKTDAFDQLVFNPNAITNQDQQRQLVDYISKRSKQFKPV
KLKHHTKFVAFSTCENFSTDNRVIVVGTIQE

SpyoM01000151 is referred to as a hypothetical protein. An example of SpyoM01000151 is shown in SEQ ID NO: 248.

SEQ ID NO: 248

MLFSVVMMLTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASF
SPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRL
VKPIPPRQPDIPKTP*LPLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyoM01000151 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 185** LPLAG (shown in italics in SEQ ID NO: 248, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyoM01000151 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in SpyoM01000151. The pilin motif sequence is underlined in SEQ ID NO: 248, below. Conserved lysine (K) residues are also marked in bold, at amino acid residue 138. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyoM01000151 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 248

MLFSVVMMLTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASF
SPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRL
VKPIPPRQPDIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

Two E boxes containing conserved glutamic residues have been identified in SpyoM01000151. The E-box motifs are underlined in SEQ ID NO: 248, below. The conserved glutamic acid (E) residues, at amino acid residues 58 and 128, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of

oligomeric pilus-like structures of SpyoM01000151. Preferred fragments of SpyoM01000151 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 248

MLFSVVMMLTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASF
SPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRL
VKPIPPRQPDIPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSRL

SpyoM01000150 is referred to as a putative MsmRL. An example of SpyoM01000150 is set forth in SEQ ID NO: 249.

SEQ ID NO: 249

MVIFDLKHVQTLHSLSQLPISVMSQDKALIQVYGNDDYLLCYYQFLKHLAIPQAAQDVIFYEGLFEESFM
IFPLCHYIIAIGPFYPYSLNKDYQEQLANNFLKHSSHRSKEELLSYMALVPHFPINNVRNLLIAIDAFFD
TQFETTCQQTIHQLLQHSKQMTADPDIIHRLKHISKASSQLPPVLEHLNHIMDLVKLGNPQLLKQEINRI
PLSSITSSSISALRAEKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEASDLLKVLRIRCAAII
HFSESLTNKSISDKRQMYNSVLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSLQHYILSTKI
KEAQLLLKRGIPVGEVAKSLYFYDTTHFHKIFKKYTGISSKDYLAKYRDNI

SpyoM01000149 is a F2 like fibronectin-binding protein. An example of SpyoM01000149 is set forth in SEQ ID NO: 250.

SEQ ID NO: 250

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGYFEIKKVDQNNKPLSGATFSLTP
KDGKGKPVQTFTSSEEGIIDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIIS
KAGSKDVSSSLQLENPKMSVVSKYGEQEKTSNSADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLD
RRLNPKGISQDIPKIIYDSENSPLAIGKYDAKTHQLTYTFTNYIAGLDKVQLSAELSLFLENKEVLENTN
ISDFKSTIGGQEITYKGTVNVLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPY
AVLNLWGFAKRTAQGENDNSSVSSAQLTGYDIYEVPHNYRLPTSYGVDISRLNLRKDLEAKLPQGSTQGA
NKRLRIDFGENLQGKAFVVKVTGKADQSGKELIVQSHLSSFNNWGSYKTLRPNSHVSFTNEIALSPSKGS
GSGTSEFTKPAITVANLKRVAQLRFKKVSTDNVPLPEAAFELRSSNGNSQKLEASSNTQGEIHFKDLTSG
TYDLYETKAPKGYQQVTEKLATVTVDTTKPAEQMVKWEKPHSFVKVEANKEVTIVNHKETLTFSGKKIWE
NDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQEYKYSVEEVKVPDGYKVSYL
GNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKKIWKNDTAENRPQAIQVQLYAD
GVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSANDIINTKREVITQQGP
NLEIEETLPLESGASGGTTTVEDSRSVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATM
ELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEIATAITFTVNEQGQVTVNGKATKGDAHIV
MVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKPSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTE
IEDSKSSDVIIGGQGQVVETTEDTQTGMHGDSGCKTEVEDTKLVQFFHFDNKEPESNSEIPKKDKPKSNT
*SLPATG*EKQHNKFFWMVTSCSLISSVFVISLKSKKRLLSC

SpyoM01000149 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 180** LPATG (shown in italics in SEQ ID NO: 250, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant SpyoM01000149 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, discussed above, containing conserved lysine (K) residues have also been identified in SpyoM01000149. The pilin motif sequences are underlined in SEQ ID NO: 250, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 157 and 163, and 216 and 224. The pilin sequences, in particular the conserved lysine residues, are thought to be important

for the formation of oligomeric, pilus-like structures. Preferred fragments of SpyoM01000149 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 250

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGYFEIKKVDQNNKPLSGATFSLTP
KDGKGKPVQTFTSSEEGIIDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIIS
KAGSKDVSSSLQLENPKMSVVSKYGEQEKTSNSADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLD
RRLNPKGISQDIPKIIYDSENSPLAIGKYDAKTHQLTYTFTNYIAGLDKVQLSAELSLFLENKEVLENTN
ISDFKSTIGGQEITYKGTVNVLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPY
AVLNLWGFAKRTAQGENDNSSVSSAQLTGYDIYEVPHNYRLPTSYGVDISRLNLRKDLEAKLPQGSTQGA
NKRLRIDFGENLQGKAFVVKVTGKADQSGKELIVQSHLSSFNNWGSYKTLRPNSHVSFTNEIALSPSKGS
GSGTSEFTKPAITVANLKRVAQLRFKKVSTDNVPLPEAAFELRSSNGNSQKLEASSNTQGEIHFKDLTSG
TYDLYETKAPKGYQQVTEKLATVTVDTTKPAEQMVKWEKPHSFVKVEANKEVTIVNHKETLTFSGKKIWE
NDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQEYKYSVEEVKVPDGYKVSYL
GNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKKIWKNDTAENRPQAIQVQLYAD
GVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSANDIINTKREVITQQGP
NLEIEETLPLESGASGGTTTVEDSRSVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATM
ELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEIATAITFTVNEQGQVTVNGKATKGDAHIV
MVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKPSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTE
IEDSKSSDVIIGGQGQVVETTEDTQTGMHGDSGCKTEVEDTKLVQFFHFDNKEPESNSEIPKKDKPKSNT
SLPATGEKQHNKFFWMVTSCSLISSVFVISLKSKKRLLSC

Two E boxes containing conserved glutamic residues have been identified in SpyoM01000149. The E-box motifs are underlined in SEQ ID NO: 250, below. The conserved glutamic acid (E) residues, at amino acid residues 329 and 668, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of SpyoM01000149. Preferred fragments of SpyoM01000149 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 250

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGYFEIKKVDQNNKPLSGATFSLTP
KDGKGKPVQTFTSSEEGIIDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIIS
KAGSKDVSSSLQLENPKMSVVSKYGEQEKTSNSADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLD
RRLNPKGISQDIPKIIYDSENSPLAIGKYDAKTHQLTYTFTNYIAGLDKVQLSAELSLFLENKEVLENTN
ISDFKSTIGGQEITYKGTVNVLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPY
AVLNLWGFAKRTAQGENDNSSVSSAQLTGYDIYEVPHNYRLPTSYGVDISRLNLRKDLEAKLPQGSTQGA
NKRLRIDFGENLQGKAFVVKVTGKADQSGKELIVQSHLSSFNNWGSYKTLRPNSHVSFTNEIALSPSKGS
GSGTSEFTKPAITVANLKRVAQLRFKKVSTDNVPLPEAAFELRSSNGNSQKLEASSNTQGEIHFKDLTSG
TYDLYETKAPKGYQQVTEKLATVTVDTTKPAEQMVKWEKPHSFVKVEANKEVTIVNHKETLTFSGKKIWE
NDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQEYKYSVEEVKVPDGYKVSYL
GNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKKIWKNDTAENRPQAIQVQLYAD
GVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSANDIINTKREVITQQGP
NLEIEETLPLESGASGGTTTVEDSRSVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDIDGKELAGATM
ELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEIATAITFTVNEQGQVTVNGKATKGDAHIV
MVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKPSDVIIGGQGEVVDTTEDTQSGMTGHSGSTTE
IEDSKSSDVIIGGQGQVVETTEDTQTGMHGDSGCKTEVEDTKLVQFFHFDNKEPESNSEIPKKDKPKSNT
SLPATGEKQHNKFFWMVTSCSLISSVFVISLKSKKRLLSC

As discussed above, applicants have also determined the nucleotide and encoded amino acid sequence of fimbrial structural subunits in several other GAS AI-3 strains of bacteria. Examples of sequences of these fimbrial structural subunits are set forth below.

M3 strain isolate ISS 3040 is a GAS AI-3 strain of bacteria. ISS3040_fimbrial is thought to be a fimbrial structural subunit of M3 strain isolate ISS 3040. An example of a nucleotide sequence

encoding the ISS3040_fimbrial protein (SEQ ID NO: 263) and an ISS3040_fimbrial protein amino acid sequence (SEQ ID NO: 264) are set forth below.

SEQ ID NO: 263

```
gagacggcaggagtgtccgaaaatgcaaaattaatagtaaaaaagacatttgactcttat
acagacaatgaagttttaatgccaaaagctgattatacttttaaagtagaggcagatagt
acagctagtggcaaaacgaaagacggtttagagattaagccaggtattgttaatggttta
acagaacagattatcagctatactaatactgataaaccagatagtaaagttaaaagtaca
gagtttgatttttcaaaagtagtattccctggtattggtgtttaccgctatactgtttca
gaaaaacaaggtgatgttgaaggaattacctacgatactaagaagtggacagtagatgtt
tatgttggaaacaaagaaggtggtggttttgaacctaagtttattgtatctaaggaacaa
ggaacagacgtcaaaaaaccagttaattttaacaactcgtttgcaactacttcgttaaaa
gttaagaagaatgtatcggggaatactggagaattgcaaaaagaatttgactttacattg
acgcttaatgaaagcacgaattttaaaaaagatcaaattgtttctttacaaaaaggaaac
gagaaatttgaagttaagattggtactccctacaagtttaaactcaaaaatggggaatct
attcaactagacaagttaccagttggtattacttataaagtcaatgaaatggaagctaat
aaagatgggtataaaacaacagcatccttgaaagagggagatggtcaatctaaaatgtat
caattggatatggaacaaaaaacagacgaatctgctgacgaaatcgttgtcacaaataag
cgtgacactcaagttccaactggtgttgtaggcacccttgctccatttgcagttcttagc
```

SEQ ID NO: 264

```
ETAGVSENAKLIVKKTFDSYTDNEVLMPKADYTFKVEADSTASG
KTKDGLEIKPGIVNGLTEQIISYTNTDKPDSKVKSTEFDFSKVVFPGIGVYRYTVSEK
QGDVEGITYDTKKWTVDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSLK
VKKNVSGNTGELQKEFDFTLTLNESTNFKKDQIVSLQKGNEKFEVKIGTPYKFKLKNG
ESIQLDKLPVGITYKVNEMEANKDGYKTTASLKEGDGQSKMYQLDMEQKTDESADEIV
VTNKRDTQVPTGVVGTLAPFAVLS
```

M44 strain isolate ISS 3776 is a GAS AI-3 strain of bacteria. ISS3776_fimbrial is thought to be a fimbrial structural subunit of M44 isolate ISS 3776. An example of a nucleotide sequence encoding the ISS3776_fimbrial protein (SEQ ID NO: 253) and an ISS3776_fimbrial protein amino acid sequence (SEQ ID NO: 254) are set forth below.

SEQ ID NO: 253

```
ttggagagagaaaaaatgaaaaaaaacaaattattacttgctactgcaatcttagcaact
gctttaggaacagcttctttaaatcaaaacgtaaaagctgagacggcaggggttgtaaca
ggaaaatcactacaagttacaaagacaatgacttatgatgatgaagaggtgttaatgccc
gaaaccgcctttacttttactatagagcctgatatgactgcaagtggaaaagaaggcagc
ctagatattaaaaatggaattgtagaaggcttagacaaacaagtaacagtaaaatataag
aatacagataaaacatctcaaaaaactaaaatagcacaatttgatttttctaaggttaaa
tttccagctataggtgtttaccgctatatggtttcagagaaaaacgataaaaaagacgga
attacgtacgatgataaaaagtggactgtagatgtttatgttgggaataaggccaataac
gaagaaggtttcgaagttctatatattgtatcaaaagaaggtacttctagtactaaaaaa
ccaattgaatttacaaactctattaaaactacttccttaaaaattgaaaaacaaataact
ggcaatgcaggagatcgtaaaaaatcattcaacttcacattaacattacaaccaagtgaa
tattataaaactggatcagttgtgaaaatcgaacaggatggaagtaaaaaagatgtgacg
ataggaacgccttacaaatttactttgggacacggtaagagtgtcatgttatcgaaatta
ccaattggtatcaattactatcttagtgaagacgaagcgaataaagacggctacactaca
acggcaacattaaaagaacaaggcaaagaaaagagttccgatttcactttgagtactcaa
aaccagaaaacagacgaatctgctgacgaaatcgttgtcacaaataagcgtgacactcaa
gttccaactggtgttgtagggacccttgctccatttgcagttcttagcattgtggctatt
ggtggagttatctatattacaaaacgtaaaaaagcttaa
```

SEQ ID NO: 254

```
MEREKMKKNKLLLATAILATALGTASLNQNVKAETAGVVTGKSL
QVTKTMTYDDEEVLMPETAFTFTIEPDMTASGKEGSLDIKNGIVEGLDKQVTVKYKNT
DKTSQKTKIAQFDFSKVKFPAIGVYRYMVSEKNDKKDGITYDDKKWTVDVYVGNKANN
EEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSLKIEKQITGNAGDRKKSFNFTLTLQP
SEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGKSVMLSKLPIGINYYLSEDEANKD
```

GVTTLATLKEQGKEKSSDFTLSTQNQKTDESADEIVVTNKRDTQVPTGVVGTLAPFAV
LSIVAIGGVIYITKRKKA

M77 strain isolate ISS4959 is a GAS AI-3 strain of bacteria. ISS4959_fimbrial is thought to be a fimbrial structural subunit of M77 strain ISS 4959. An example of a nucleotide sequence encoding the ISS4959_fimbrial protein (SEQ ID NO: 271) and an ISS4959_fimbrial protein amino acid sequence (SEQ ID NO: 272) are set forth below.

SEQ ID NO: 271

gtaacagtaaaatataagaatacagataaaacatctcaaaaaactaaaatagcacaattt
gatttttctaaggttaaatttccagctataggtgtttaccgctatatggtttcagagaaa
aacgataaaaaagacggaattacgtacgatgataaaaagtggacngtagatgtttatgtt
gggaataaggccaataacgaagaaggtttcgaagttctatatattgtatcaaaagaaggt
acttctagtnctaaaaaaccaattgaatttacaaactctattaaaactacttccttaaaa
attgaaaaacaaataactggcaatgcaggagatcgtaaaaaatcattcaacttcacattn
acattacanccaagtgaatattataaaactggatcagttgtgaaaatcgaacaggatgga
agtaaaaaagatgtgacgataggaacgccttacaaatttactttgggacacggtaagagt
gtcatgttatcgaaattnccaattggtatcaattactatcttagtgaagacgaagcgaat
aaagacggntacactacancggcaacattaaaagaacaaggcaaagaaaagagttccgat
ttcactttgagtactcaaaaccagaaaacagacgaatctgctg

SEQ ID NO: 272

VTVKYKNTDKTSQKTKIAQFDFSKVKFPAIGVYRYMVSEKNDKK
DGITYDDKKWTVDVYVGNKANNEEGFEVLYIVSKEGTSSXKKPIEFTNSIKTTSLKIE
KQITGNAGDRKKSFNFTXTLXPSEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGKS
VMLSKXPIGINYYLSEDEANKDGYTTXATLKEQGKEKSSDFTLSTQNQKTDESA

Examples of GAS AI-4 sequences from M12 strain isolate A735 are set forth below.

19224133 is thought to be a RofA regulatory protein. An example of a nucleotide sequence encoding the RofA regulatory protein (SEQ ID NO: 104) and a RofA regulatory protein amino acid sequence (SEQ ID NO: 105) are set forth below.

SEQ ID NO: 104

ATGACCATCCAAAAAAGGATGATATCTTGCCAATTTACACATCCTTCTAAAGAAACTTATCTTTACCAACTCTAT
GCATCATCTAATGTCTTACAATTACTAGCGTTTTTAATAAAAAATGGTTCCCACTCTCGTCCCCTTACGGATTTT
GCAAGAAGTCATTTTTTATCAAACTCCTCAGCTTATCGGATGCGCGAAGCATTGATTCCTTTATTAAGAAACTTT
GAATTAAAACTCTCTAAGAACAAGATTGTCGGTGAGGAATATCGTATCCGTTACCTCATCGCTCTGCTATATAGT
AAGTTTGGCATTAAAGTTTATGACTTGACGCAGCAAGACAAAAACATTATTCATAGCTTTTTATCCCATAGTTCC
ACCCACCTTAAAACTTCTCCTTGGTTATCGGAATCGTTTTCTTTCTATGACATTTTATTAGCTTTATCGTGGAAG
CGGCATCAATTTTCGGTAACTATTCCCCAAACCAGAATTTTTCAACAATTAAAAAAACTTTTTGTCTACGATTCT
TTGAAAAAAAGTAGCCGTGATATTATCGAAACTTACTGCCAACTAAACTTTTCAGCAGGAGATTTGGACTACCTC
TATTTAATTTATATCACCGCTAATAATTCTTTTGCGAGCTTACAATGGACACCTGAGCATATCAGACAATGTTGT
CAACTTTTTGAAGAAAATGATACTTTTCGCCTGCTTTTAAATCCTATCATCACTCTTTTACCTAACCTAAAAGAG
CAAAAGGCTAGTTTAGTAAAAGCTCTTATGTTTTTTTCAAAATCATTCTTGTTTAATCTGCAACATTTTATTCCT
GAGACCAACTTATTCGTTTCTCCGTACTATAAAGGAAACCAAAAACTCTATACGTCCTTAAAGTTAATTGTCGAA
GAGTGGATGGCCAAACTTCCTGGTAAGCGTTACTTGAACCATAAGCATTTTCATCTTTTTTGCCACTATGTCGAG
CAAATTCTAAGAAATATCCAACCTCCTTTAGTTGTTGTTTTCGTAGCCAGTAATTTTATCAATGCTCATCTCCTA
ACAGATTCTTTCCCAAGGTATTTCTCGGATAAAAGCATTGATTTTCATTCCTATTATCTATTGCAAGATAATGTT
TATCAAATTCCTGATTTAAAGCCAGATTTGGTCATCACTCACAGTCAACTGATTCCTTTTGTTCACCATGAACTT
ACAAAAGGAATTGCTGTTGCTGAAATATCTTTTGATGAATCGATTCTGTCTATCCAAGAATTGATGTATCAAGTT
AAAGAGGAAAAATTCCAAGCTGATTTAACCAAACAATTAACATAA

SEQ ID NO: 105

MTIQKRMISCQFTHPSKETYLYQLYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNSSAYRMREALIPLLRNF
ELKLSKNKIVGEEYRIRYLIALLYSKFGIKVYDLTQQDKNIIHSFLSHSSTHLKTSPWLSESFSFYDILLALSWK
RHQFSVTIPQTRIFQQLKKLFVYDSLKKSSRDIIETYCQLNFSAGDLDYLYLIYITANNSFASLQWTPEHIRQCC
QLFEENDTFRLLLNPIITLLPNLKEQKASLVKALMFFSKSFLFNLQHFIPETNLFVSPYYKGNQKLYTSLKLIVE

EWMAKLPGKRYLNNHKHFHLFCHYVEQTLRNTQPPLVVVFVASNFINAHLLTDSFPRYFSDKSIDFHSYYLLQDNV
YQIPDLKPDLVITHSQLIPFVHHELTKGIAVAEISFDESILSIQELMYQVKEEKFQADLTKQLT

19224134 is thought to be a protein F fibronectin binding protein. An example of a nucleotide sequence encoding the protein F fibronectin binding protein (SEQ ID NO: 106) and a protein F fibronectin binding protein amino acid sequence (SEQ ID NO: 107) are set forth below.

SEQ ID NO: 106

ATGGTAAGCTCATATATGTTTGCGAGAGGAGAGAAAATGAATAACAAAATGTTTTTGAACAAAGAAGCCGGTTTT
TTGGTACACACAAAAAGAAAAAGGCGATTTGCTGTCACTTTAGTGGGAGTCTTTTTTCTGCTTTTGGCATGTGCG
GGTGCTATCGGTTTTGGTCAAGTAGCCTATGCTGCGGATGAGAAGACTGTGCCGAATTTTAAAAGCCCAGATCCA
GATTATCCCTGGTATGGTTATGATTCGTATAGAGGAATATTTGCAAGATATCACAATTTAAAAGTAAATCTAAAA
GGAAGTAAGGAGTATCAAGCGTATTGTTTTAACCTAACAAAATACTTTCCTCGCCCCACTTATAGTACTACAAAT
AATTTTTACAAGAAAATTGATGGGAGTGGATCAGCGTTCAAATCTTATGCAGCGAATCCTAGGGTTTTAGATGAG
AATTTAGATAAATTAGAAAAAAATATACTGAATGTAATTTATAATGGATATAAAAGTAATGCAAATGGTTTTATG
AATGGTATAGAAGATCTTAATGCTATACTAGTAACTCAAAACGCTATTTGGTACTATTCAGATAGTGCTCCATTA
AATGATGTTAATAAAATGTGGGAAAGAGAGGTTCGGAATGGGGAGATTAGTGAGTCACAAGTTACTTTAATGCGT
GAGGCATTGAAAAAACTAATTGATCCCAATTTAGAAGCTACTGCAGCTAATAAAATCCCATCAGGATATCGTTTA
AATATCTTTAAGTCTGAAAATGAAGATTACCAAAATCTTTTAAGTGCTGAATATGTACCTGATGATCCCCCTAAA
CCTGGTGATACGTCAGAACATAATCCTAAAACTCCCGAGTTGGATGGCACTCCAATTCCCGAGGACCCAAAACGT
CCAGATGAGAGTTCAGAACCTGCGCTTCCCCCATTAATGCCAGAGCTAGATGGTGAAGAAGTCCCAGAAGTTCCA
AGCGAGAGCTTAGAACCTGCGCTTCCCCCATTGATGCCAGAGCTAGATGGTGAAGAAGTCCCAGAAGTTCCAAGC
GAGAGCTTAGAACCTGCGCTTCCCCCATTGATGCCAGAGCTAGATGGTGAAGAAGTCCCAGAAGTTCCAAGCGAG
AGCTTAGAACCTGCGCTTCCCCCATTAATGCCAGAGCTAGATGGTGAAGAAGTCCCAGAAGTTCCAAGCGAGAGC
TTAGAACCTGCGCTTCCCCCATTGATGCCAGAGTTAGATGGTGAAGAAGTCCCTGAAAAACCTAGTGTTGACTTA
CCTATTGAAGTTCCTCGTTATGAGTTTAACAATAAAGACCAGTCACCTCTAGCGGGTGAGTCTGGTGAGACGGAG
TATATTACCGAAGTCTATGGAAATCAACAGAACCCTGTTGATATTGATAAAAAACTTCCGAATGAAACAGGTTTT
TCAGGAAATATGGTTGAGACAGAAGATACGAAAGAGCCAGAAGTGTTGATGGGAGGTCAAAGTGAGTCTGTTGAA
TTTACTAAAGACACTCAAACAGGCATGAGTGGTCAAACAACTCCTCAGGTTGAGACAGAAGATACGAAAGAGCCA
GAAGTGTTGATGGGAGGTCAAAGTGAGTCTGTTGAATTTACTAAAGACACTCAAACAGGCATGAGTGGTCAAACA
ACTCCTCAGGTTGAGACAGAAGATACGAAAGAGCCAGGAGTGTTGATGGGAGGCCAAAGTGAGTCTGTTGAATTT
ACTAAAGACACTCAAACAGGCATGAGTGGTCAAACAACTCCTCAGGTTGAGACAGAAGACACGAAAGAGCCAGGA
GTGTTGATGGGAGGTCAAAGTGAGTCTGTTGAATTTACTAAAGACACTCAAACAGGCATGAGCGGTTTCAGTGAA
ACAGTGACCATTGTTGAAGATACGCGTCCGAAGTTAGTGTTCCATTTTGACAATAATGAGCCCAAAGTGGAAGAG
AATCGGGAAAAGCCTACAAAAAATATAACACCTATCCTTCCTGCAACAGGAGATATTGAGAATGTTTTGGCCTTT
CTTGGAATCCTTATTTTGTCAGTACTTTCTATTTTTAGCCTTTTAAAAAACAAACAAAACAATAAAGTCTGA

SEQ ID NO: 107

MVSSYMFARGEKMNNKMFLNKEAGFLVHTKRKRRFAVTLVGVFFLLLACAGAIGFGQVAYAADEKTVPNFKSPDP
DYPWYGYDSYRGIFARYHNLKVNLKGSKEYQAYCFNLTKYFPRPTYSTTNNFYKKIDGSGSAFKSYAANPRVLDE
NLDKLEKNILNVIYNGYKSNANGFMNGIEDLNAILVTQNAIWYYSDSAPLNDVNKMWEREVRNGEISESQVTLMR
EALKKLIDPNLEATAANKIPSGYRLNIFKSENEDYQNLLSAEYVPDDPPKPGDTSEHNPKTPELDGTPIPEDPKR
PDESSEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSE
SLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEKPSVDLPIEVPRYEFNNKDQSPLAGESGETE
YITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEP
EVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPGVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPG
VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDNNEPKVEENREKPTKNITPI*LPATG*DIENVLAF
LGILILSVLSIFSLLKNKQNNKV

19224134 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 181** LPATG (shown in italics in SEQ ID NO: 107, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 19224134 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular

domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in 19224134. The pilin motif sequence is underlined in SEQ ID NO: 107, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 275, 285, and 299. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of 19224134 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 107

MVSSYMFARGEKMNNKMFLNKEAGFLVHTKRKRRFAVTLVGVFFLLLACAGAIGFGQVAYAADEKTVPNFKSPDP
DYPWYGYDSYRGIFARYHNLKVNLKGSKEYQAYCFNLTKYFPRPTYSTTNNFYKKIDGSGSAFKSYAANPRVLDE
NLDKLEKNILNVIYNGYKSNANGFMNGIEDLNAILVTQNAIWYYSDSAPLNDVNKMWEREVRNGEISESQVTLMR
EALKKLIDPNLEATAANKIPSGYRLNIFKSENEDYQNLLSAEYVPDDPPKPGDTSEHNPKTPELDGTPIPEDPKR
PDESSEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSE
SLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEKPSVDLPIEVPRYEFNNKDQSPLAGESGETE
YITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEP
EVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPGVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPG
VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDNNEPKVEENREKPTKNITPILPATGDIENVLAF
LGILILSVLSIFSLLKNKQNNKV

Two E boxes containing conserved glutamic residues have been identified in 19224134. The E-box motifs are underlined in SEQ ID NO: 107, below. The conserved glutamic acid (E) residues, at amino acid residues 487 and 524, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of 19224134. Preferred fragments of 19224134 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 107

MVSSYMFARGEKMNNKMFLNKEAGFLVHTKRKRRFAVTLVGVFFLLLACAGAIGFGQVAYAADEKTVPNFKSPDP
DYPWYGYDSYRGIFARYHNLKVNLKGSKEYQAYCFNLTKYFPRPTYSTTNNFYKKIDGSGSAFKSYAANPRVLDE
NLDKLEKNILNVIYNGYKSNANGFMNGIEDLNAILVTQNAIWYYSDSAPLNDVNKMWEREVRNGEISESQVTLMR
EALKKLIDPNLEATAANKIPSGYRLNIFKSENEDYQNLLSAEYVPDDPPKPGDTSEHNPKTPELDGTPIPEDPKR
PDESSEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEVPSE
SLEPALPPLMPELDGEEVPEVPSESLEPALPPLMPELDGEEVPEKPSVDLPIEVPRYEFNNKDQSPLAGESGETE
YITEVYGNQQNPVDIDKKLPNETGFSGNMVETEDTKEPEVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEP
EVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPGVLMGGQSESVEFTKDTQTGMSGQTTPQVETEDTKEPG
VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDNNEPKVEENREKPTKNITPILPATGDIENVLAF
LGILILSVLSIFSLLKNKQNNKV

19224135 is thought to be a capsular polysaccharide adhesin (Cpa) protein. An example of a nucleotide sequence encoding the Cpa protein (SEQ ID NO: 108) and a Cpa protein amino acid sequence (SEQ ID NO: 109) are set forth below.

SEQ ID NO: 108

ATGAATAACAAAAAATTGCAAAAGAAGCAAGATGCTCCTCGGGTATCAAACAGAAAGCCAAAACAATTAACTGTC
ACTTTAGTGGGAGTATTTTTAATGTTTTTGACCTTGGTAAGTTCCATGAGAGGTGCTCAAAGCATATTTGGAGAG
GAAAAGAGAATTGAAGAAGTCAGTGTTCCTAAAATAAAAAGTCCAGATGATGCCTACCCTTGGTATGGCTATGAT
TCATATGACTCTAGTCATCCTTACTATGAACGTTTTAAAGTAGCACATGATTTAAGGGTTAATTTAAATGGAAGT
AAGAGCTACCAAGTATATTGCTTTAATATCAATTCTCATTATCCGAATAGAAAAAATGCTTTTTCTAAACAATGG
TTTAAGAGAGTTGATGGGACAGGTGATGTGTTCACAAATTATGCTCAGACACCTAAGATTCGTGGAGAATCATTG
AATAATAAACTTTTAAGTATTATGTACAACGCTTATCCTAAAAATGCTAATGGCTATATGGATAAGATAGAACCA

TTAAAATGCTATTTTAGTAACTCAACAAGCTGTTTGGTACTATTCTGACAGTTCTTATGGTAATATAAAAACGTTA
TGGGCATCTGAGCTTAAAGACGGAAAAATAGATTTTGAACAAGTAAAATTAATGCGTGAAGCTTACTCAAAACTA
ATTAGTGATGATTTAGAAGAAACATCTAAAAATAAGCTACCTCAAGGATCTAAACTGAATATTTTTGTTCCGCAA
GATAAATCTGTTCAAAATTTATTAAGTGCAGAGTACGTGCCTGAATCCCCTCCGGCACCAGGTCAGTCTCCAGAA
CCGCCAGTGCAAACAAAAAAAACATCAGTCATTATCAGAAAATATGCGGAAGGTGACTACTCTAAACTTCTAGAG
GGAGCAACTTTGCGTTTAACAGGGGAAGATATCCTAGATTTTCAAGAAAAAGTCTTCCAAAGTAATGGAACAGGA
GAAAAGATTGAATTATCAAATGGGACTTATACCTTAACAGAAACATCATCTCCAGATGGATATAAAATTGCGGAG
CCGATTAAGTTTAGAGTAGTGAATAAAAAAGTATTTATCGTCCAAAAAGATGGTTCTCAAGTGGAAAATCCAAAC
AAAGAAGTAGCAGAGCCATACTCAGTGGAAGCGTACAGCGATATGCAAGATAGTAACTATATTAATCCAGAAACG
TTCACTCCTTATGGGAAATTTTATTACGCTAAAAATAAGGATAAAAGTTCACAAGTTGTCTACTGTTTTAATGCT
GATTTACACTCTCCACCTGAATCAGAGGATGGGGGAGGAACTATAGATCCTGATATTAGTACGATGAAAGAAGTC
AAGTACACACATACGGCAGGTAGTGATTTGTTTAAATACGCGCTAAGACCGAGAGATACAAATCCAGAAGACTTC
TTAAAGCACATTAAAAAAGTAATTGAAAAAGGCTACAATAAAAAAGGTGATAGCTATAATGGATTAACAGAAACA
CAGTTTCGCGCGGCTACTCAGCTTGCTATCTATTACTTTACAGACAGCACTGACTTAAAAACCTTAAAAACTTAT
AACAATGGGAAAGGTTACCATGGATTTGAATCTATGGATGAAAAAACCCTAGCTGTAACAAAAGAATTAATTAAT
TACGCTCAAGATAATAGTGCCCCTCAACTAACAAATCTTGATTTCTTCGTACCTAATAATAGCAAATACCAATCT
CTTATTGGGACAGAATACCATCCAGATGATTTGGTTGACGTGATTCGTATGGAAGATAAAAAGCAAGAAGTTATT
CCAGTAACTCACAGTTTGACAGTGAAAAAAACAGTAGTCGGTGAGTTGGGAGATAAAACTAAAGGCTTCCAATTT
GAACTTGAGTTGAAAGATAAAACTGGACAGCCTATTGTTAACACTCTAAAAACTAATAATCAAGATTTAGTAGCT
AAAGATGGGAAATATTCATTTAATCTAAAGCATGGTGACACCATAAGAATAGAAGGATTACCGACGGGATATTCT
TATACTCTGAAAGAGACTGAAGCTAAGGATTATATAGTAACCGTTGATAACAAAGTTAGTCAAGAAGCTCAATCA
GCAAGTGAGAATGTCACAGCAGACAAAGAAGTCACTTTTGAAAACCGTAAAGATCTTGTCCCACCAACTGGTTTT
ATTACTGATGGTGGAACCTATCTGTGGTTATTATTGCTTGTCCCATTTGGTTTGTTAGTGTGGTTCTTTGGTCGT
AAAGGACTAAAAAATGACTAA

SEQ ID NO: 109

MNNKKLQKKQDAPRVSNRKPKQLTVTLVGVFLMFLTLVSSMRGAQSIFGEEKRIEEVSVPKIKSPDDAYPWYGYD
SYDSSHPYYERFKVAHDLRVNLNGSKSYQVYCFNINSHYPNRKNAFSKQWFKRVDGTGDVFTNYAQTPKIRGESL
NNKLLSIMYNAYPKNANGYMDKIEPLNAILVTQQAVWYYSDSSYGNIKTLWASELKDGKIDFEQVKLMREAYSKL
ISDDLEETSKNKLPQGSKLNIFVPQDKSVQNLLSAEYVPESPPAPGQSPEPPVQTKKTSVIIRKYAEGDYSKLLE
GATLRLTGEDILDFQEKVFQSNGTGEKIELSNGTYTLTETSSPDGYKIAEPIKFRVVNKKVFIVQKDGSQVENPN
KEVAEPYSVEAYSDMQDSNYINPETFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPESEDGGGTIDPDISTMKEV
KYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYNKKGDSYNGLTETQFRAATQLAIYYFTDSTDLKTLKTY
NNGKGYHGFESMDEKTLAVTKELINYAQDNSAPQLTNLDFFVPNNSKYQSLIGTEYHPDDLVDVIRMEDKKQEVI
PVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYS
YTLKETEAKDYIVTVDNKVSQEAQSASENVTADKEVTFENRKDL*VPPTG*FITDGGTYLWLLLLVPFGLLVWFFGR
KGLKND

19224135 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 184** VPPTG (shown in italics in SEQ ID NO: 109, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 19224135 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in 19224135. The pilin motif sequence is underlined in SEQ ID NO: 109, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 164 and 172. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of 19224135 include at least one conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 109

MNNKKLQKKQDAPRVSNRKPKQLTVTLVGVFLMFLTLVSSMRGAQSIFGEEKRIEEVSVPKIKSPDDAYPWYGYD
SYDSSHPYYERFKVAHDLRVNLNGSKSYQVYCFNINSHYPNRKNAFSKQWFKRVDGTGDVFTNYAQTPKIRGESL
NNKLLSIMYNAYPKNANGYMDKIEPLNAILVTQQAVWYYSDSSYGNIKTLWASELKDGKIDFEQVKLMREAYSKL
ISDDLEETSKNKLPQGSKLNIFVPQDKSVQNLLSAEYVPESPPAPGQSPEPPVQTKKTSVIIRKYAEGDYSKLLE
GATLRLTGEDILDFQEKVFQSNGTGEKIELSNGTYTLTETSSPDGYKIAEPIKFRVVNKKVFIVQKDGSQVENPN
KEVAEPYSVEAYSDMQDSNYINPETFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPESEDGGGTIDPDISTMKEV
KYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYNKKGDSYNGLTETQFRAATQLAIYYFTDSTDLKTLKTY
NNGKGYHGFESMDEKTLAVTKELINYAQDNSAPQLTNLDFFVPNNSKYQSLIGTEYHPDDLVDVIRMEDKKQEVI
PVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYS
YTLKETEAKDYIVTVDNKVSQEAQSASENVTADKEVTFENRKDLVPPTGFITDGGTYLWLLLLVPFGLLVWFFGR
KGLKND

An E box containing a conserved glutamic residue has been identified in 19224135. The E-box motif is underlined in SEQ ID NO: 109, below. The conserved glutamic acid (E), at amino acid residue 339, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of 19224135. Preferred fragments of 19224135 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 109

MNNKKLQKKQDAPRVSNRKPKQLTVTLVGVFLMFLTLVSSMRGAQSIFGEEKRIEEVSVPKIKSPDDAYPWYGYD
SYDSSHPYYERFKVAHDLRVNLNGSKSYQVYCFNINSHYPNRKNAFSKQWFKRVDGTGDVFTNYAQTPKIRGESL
NNKLLSIMYNAYPKNANGYMDKIEPLNAILVTQQAVWYYSDSSYGNIKTLWASELKDGKIDFEQVKLMREAYSKL
ISDDLEETSKNKLPQGSKLNIFVPQDKSVQNLLSAEYVPESPPAPGQSPEPPVQTKKTSVIIRKYAEGDYSKLLE
GATLRLTGEDILDFQEKVFQSNGTGEKIELSNGTYTLTETSSPDGYKIAEPIKFRVVNKKVFIVQKDGSQVENPN
KEVAEPYSVEAYSDMQDSNYINPETFTPYGKFYYAKNKDKSSQVVYCFNADLHSPPESEDGGGTIDPDISTMKEV
KYTHTAGSDLFKYALRPRDTNPEDFLKHIKKVIEKGYNKKGDSYNGLTETQFRAATQLAIYYFTDSTDLKTLKTY
NNGKGYHGFESMDEKTLAVTKELINYAQDNSAPQLTNLDFFVPNNSKYQSLIGTEYHPDDLVDVIRMEDKKQEVI
PVTHSLTVKKTVVGELGDKTKGFQFELELKDKTGQPIVNTLKTNNQDLVAKDGKYSFNLKHGDTIRIEGLPTGYS
YTLKETEAKDYIVTVDNKVSQEAQSASENVTADKEVTFENRKDLVPPTGFITDGGTYLWLLLLVPFGLLVWFFGR
KGLKND

19224136 is thought to be a LepA protein. An example of a nucleotide sequence encoding the LepA protein (SEQ ID NO: 110) and a LepA protein amino acid sequence (SEQ ID NO: 111) are set forth below.

SEQ ID NO: 110

ATGACTAATTACCTAAATCGCTTAAATGAGAATCCACTATTTAAAGCTTTCATACGGTTAGTACTTAAGATTTCT
ATTATTGGATTTCTAGGTTACATTCTATTTCAGTATGTTTTTGGCGTCATGATTGTTAACACAAATCAGATGAGT
CCTGCTGTAAGTGCTGGTGATGGAGTCTTATATTATCGTTTGACTGATCGCTATCATATTAATGATGTGGTGGTC
TATGAGGTTGATAACACTTTGAAAGTTGGTCGAATTGCCGCTCAAGCTGGCGATGAGGTTAGTTTTACGCAAGAA
GGAGGACTGTTGATTAATGGGCATCCACCAGAAAAAGAGGTCCCTTACCTGACGTATCCTCACTCAAGTGGTCCA
AACTTTCCCTATAAAGTTCCTACGGGTACGTATTTCATATTGAATGATTATCGTGAAGAACGTTTGGACAGTCGT
TATTATGGGGCGTTACCCATCAATCAAATCAAAGGGAAAATCTCAACTCTATTAAGAGTGAGAGGAATTTAA

SEQ ID NO: 111

MTNYLNRLNENPLFKAFIRLVLKISIIGFLGYILFQYVFGVMIVNTNQMSPAVSAGDGVLYYRLTDRYHINDVVV
YEVDNTLKVGRIAAQAGDEVSFTQEGGLLINGHPPEKEVPYLTYPHSSGPNFPYKVPTGTYFILNDYREERLDSR
YYGALPINQIKGKISTLLRVRGI

19224137 is thought to be a fimbrial protein. An example of a nucleotide sequence encoding the fimbrial protein (SEQ ID NO: 112) and a fimbrial protein amino acid sequence (SEQ ID NO: 113) are set forth below.

SEQ ID NO: 112

ATGAAAAAAAATAAATTATTACTTGCTACTGCAATCTTAGCAACTGCTTTAGGAACAGCTTCTTTAAATCAAAAC
GTAAAAGCTGAGACGGCAGGGGTTGTTAGCAGTGGTCAATTAACAATAAAAAAATCAATTACAAATTTTAATGAT
GATACACTTTTGATGCCTAAGACAGACTATACTTTTAGCGTTAATCCGGATAGTGCGGCTACAGGTACTGAAAGT
AATTTACCAATTAAACCAGGTATTGCTGTTAACAATCAAGATATTAAGGTTTCTTATTCTAATACTGATAAGACA
TCAGGTAAAGAAAAACAAGTTGTTGTTGACTTTATGAAAGTTACTTTTCCTAGCGTTGGTATTTACCGTTATGTT
GTTACCGAGAATAAAGGGACAGCAGAAGGAGTTACATATGATGATACAAAATGGTTAGTTGACGTCTATGTTGGT
AATAATGAAAAGGGAGGTCTTGAACCAAAGTATATTGTATCTAAAAAAGGAGATTCTGCTACTAAAGAACCAATC
CAGTTTAATAATTCATTCGAAACAACGTCATTAAAAATTGAAAAGGAAGTTACTGGTAATACAGGAGATCATAAA
AAAGCATTTACCTTTACATTAACATTGCAACCAAATGAATACTATGAGGCAAGTTCGGTTGTGAAAATTGAAGAG
AACGGACAAACGAAAGATGTGAAAATTGGGGAGGCATATAAGTTTACTTTGAACGATAGTCAGAGTGTGATATTG
TCTAAATTACCAGTTGGTATTAATTATAAAGTTGAAGAAGCAGAAGCTAATCAAGGTGGATATACTACAACAGCA
ACTTTAAAAGATGGAGAAAAGTTATCTACTTATAACTTAGGTCAGGAACATAAAACAGACAAGACTGCTGATGAA
ATCGTTGTCACAAATAACCGTGACACTCAAGTTCCAACTGGTGTTGTAGGCACCCTTGCTCCATTTGCAGTTCTT
AGCATTGTGGCTATTGGTGGAGTTATCTATATTACAAAACGTAAAAAAGCTTAA

SEQ ID NO: 113

MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATGTES
NLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENKGTAEGVTYDDTKWLVDVYVG
NNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKIEKEVTGNTGDHKKAFTFTLTLQPNEYYEASSVVKIEE
NGQTKDVKIGEAYKFTLNDSQSVILSKLPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADE
IVVTNNRDT*QVPTG*VVGTLAPFAVLSIVAIGGVIYITKRKKA

19224137 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 140** QVPTG (shown in italics in SEQ ID NO: 113, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 19224137 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in 19224137. The pilin motif sequence is underlined in SEQ ID NO: 113, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 160. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of 19224137 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 113

MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATGTES
NLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENKGTAEGVTYDDTKWLVDVYVG
NNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKIEKEVTGNTGDHKKAFTFTLTLQPNEYYEASSVVKIEE
NGQTKDVKIGEAYKFTLNDSQSVILSKLPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADE
IVVTNNRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

An E box containing a conserved glutamic residue has been identified in 19224137. The E-box motif is underlined in SEQ ID NO: 113, below. The conserved glutamic acid (E), at amino acid residue 263, is marked in bold. The E box motif, in particular the conserved glutamic acid residue, is thought to be important for the formation of oligomeric pilus-like structures of 19224137. Preferred fragments of 19224137 include the conserved glutamic acid residue. Preferably, fragments include the E box motif.

SEQ ID NO: 113

MKKNKLLATALATAIGTASINQVKAETAGVVSSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATGTES
NLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENKGTAEGVTYDDTKWLVDVYVG
NNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKIEKEVTGNTGDHKKAFTFTLTLQPNEYYEASSVVKIEE
NGQTKDVKIGEAYKFTLNDSQSVILSKLPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADE
IVVTNNRDTQVPTGVVGTLAPFAVLSIVAIGGVIYITKRKKA

19224138 is thought to be a SrtC2-type sortase. An example of a nucleotide sequence encoding the SrtC2 sortase (SEQ ID NO: 114) and a SrtC2 sortase amino acid sequence (SEQ ID NO: 115) are set forth below.

SEQ ID NO: 114

ATGATGATGACAATTGTACAGGTTATCAATAAAGCCATTGATACTCTCATTCTTATCTTTTGTTTAGTCGTACTA
TTTTTAGCTGGTTTTGGTTTGTGGGATTCTTATCATCTCTATCAACAAGCAGACGCTTCTAATTTCAAAAAATTT
AAAACAGCTCAACAACAGCCTAAATTTGAAGACTTGTTAGCTTTGAATGAGGATGTCATTGGTTGGTTAAATATC
CCGGGGACTCATATTGATTATCCTCTAGTTCAGGGAAAAACGAATTTAGAGTATATTAATAAAGCAGTTGATGGC
AGTGTTGCCATGTCTGGTAGTTTATTTTTAGATACACGGAATCATAATGATTTTACGGACGATTACTCTCTGATT
TATGGCCATCATATGGCAGGTAATGCCATGTTTGGCGAAATTCCAAAATTTTTAAAAAAGGATTTTTTCAACAAA
CATAATAAAGCTATCATTGAAACAAAAGAGAGAAAAAAACTAACCGTCACTATTTTTGCTTGTCTCAAGACAGAT
GCCTTTGACCAGTTAGTTTTTAATCCTAATGCTATTACCAATCAAGACCAACAAAGGCAGCTCGTTGATTATATC
AGTAAAAGATCAAAACAATTTAAACCTGTTAAATTGAAGCATCATACAAAGTTCGTTGCTTTTTCAACGTGTGAA
AATTTTTCTACTGACAATCGTGTTATCGTTGTCGGTACTATTCAAGAATAA

SEQ ID NO: 115

MMMTIVQVINKAIDTLILIFCLVVLFLAGFGLWDSYHLYQQADASNFKKFKTAQQQPKFEDLLALNEDVIGWLNI
PGTHIDYPLVQGKTNLEYINKAVDGSVAMSGSLFLDTRNHNDFTDDYSLIYGHHMAGNAMFGEIPKFLKKDFFNK
HNKAIIETKERKKLTVTIFACLKTDAFDQLVFNPNAITNQDQQRQLVDYISKRSKQFKPVKLKHHTKFVAFSTCE
NFSTDNRVIVVGTIQE

19224139 is an open reading frame that encodes a sortase substrate motif LPXAG shown in italics in SEQ ID NO: 117. An example of a nucleotide sequence of the open reading frame (SEQ ID NO: 116) and the amino acid sequence encoded by the open reading frame (SEQ ID NO: 117) are set forth below.

SEQ ID NO: 116

ATGTTATTTTCTGTCGTAATGATATTAACCATGCTGGCCTTTAATCAGACTGTTTTAGCAAAAGACAGCACTGTT
CAAACTAGCATTAGTGTCGAAAATGTCTTAGAGAGAGCAGGCGATAGTACCCCATTTTCGATTGCATTAGAATCA
ATTGATGCGATGAAAACAATAGAAGAAATAACAATTGCTGGTTCTGGAAAAGCAAGCTTTTCCCCTCTGACCTTC
ACAACAGTTGGGCAATATACTTATCGTGTTTATCAGAAGCCTTCACAAAATAAAGATTATCAAGCAGATACTACT
GTATTTGACGTTCTTGTCTATGTGACCTATGATGAAGATGGGACTCTAGTCGCAAAAGTTATTTCTCGAAGGGCT
GGAGACGAAGAAAAATCAGCGATTACTTTTAAGCCCAAACGGTTAGTAAAACCAATACCGCCTAGACAACCTAAC
ATCCCTAAAACCCCATTACCATTAGCTGGTGAAGTAAAAAGTTTATTGGGTATCTTAAGTATCGTATTACTGGGG
TTACTAGTTCTTCTTTATGTTAAAAAACTGAAGAG

SEQ ID NO: 117

MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTF
TTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPN
IPKTP*LPLAG*EVKSLLGILSIVLLGLLVLLYVKKLKSKL

19224139 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 185** LPLAG (shown in italics in SEQ ID NO: 117, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 19224139 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular

domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

A pilin motif, discussed above, containing a conserved lysine (K) residue has also been identified in 19224139. The pilin motif sequence is underlined in SEQ ID NO: 117, below. A conserved lysine (K) residue is also marked in bold, at amino acid residue 138. The pilin sequence, in particular the conserved lysine residue, is thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of 19224139 include the conserved lysine residue. Preferably, fragments include the pilin sequence.

SEQ ID NO: 117

MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTF
TTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPN
IPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSKL

Two E boxes containing conserved glutamic residues have been identified in 19224139. The E-box motifs are underlined in SEQ ID NO: 117, below. The conserved glutamic acid (E) residues, at amino acid residues 58 and 128, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of 19224139. Preferred fragments of 19224139 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 117

MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTPFSIALESIDAMKTIEEITIAGSGKASFSPLTF
TTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVTYDEDGTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPN
IPKTPLPLAGEVKSLLGILSIVLLGLLVLLYVKKLKSKL

19224140 is thought to be a MsmRL protein. An example of a nucleotide sequence encoding the MsmRL protein (SEQ ID NO: 118) and a MsmRL protein amino acid sequence (SEQ ID NO: 119) are set forth below.

SEQ ID NO: 118

ATGGTTATATTCGATTTAAAACATGTGCAAACATTACACAGCTTGTCTCAATTACCTATTTCAGTGATGTCACAA
GATAAGGCACTTATTCAAGTATATGGTAATGACGACTATTTATTATGTTACTATCAATTTTTAAAGCATCTAGCT
ATTCCTCAAGCTGCACAAGATGTTATTTTTTATGAGGGTTTATTTGAAGAGTCCTTTATGATTTTTCCTCTTTGT
CACTACATTATTGCCATTGGACCTTTCTACCCTTATTCACTTAATAAAGACTATCAGGAACAATTAGCTAATAAT
TTTTTAAAACATTCTTCTCATCGTAGCAAAGAAGAGCTCTTATCCTATATGGCACTTGTCCCACATTTTCCAATT
AATAATGTGCGGAACCTTTTGATAGCTATTGACGCTTTTTTTGACACACAATTTGAGACGACTTGCCAACAAACA
ATTCATCAATTGTTGCAGCATTCAAAACAGATGACTGCTGATCCTGATATCATTCATCGCCTTAAGCATATTAGC
AAAGCATCTAGCCAACTACCGCCTGTTTTAGAGCACCTAAATCATATTATGGATCTGGTAAAGCTAGGCAATCCA
CAATTGCTCAAGCAAGAAATCAATCGCATCCCCTTATCAAGTATCACCTCATCTTCTATTTCTGCTCTAAGGGCG
GAAAAGAACCTCACTGTTATCTATTTAACTAGGTTACTGGAATTCAGTTTTGTAGAAAATACTGACGTAGCAAAG
CATTATAGCCTTGTCAAATACTACATGGCCTTAAATGAAGAAGCGAGTGACTTGCTCAAAGTTTTGAGAATTCGC
TGTGCAGCCATCATCCATTTTTCCGAATCATTAACCAATAAAAGTATTTCTGATAAACGTCAAATGTACAATAGT
GTGCTTCATTATGTCGATAGTCACCTGTATTCCAAATTAAAGGTATCTGATATCGCTAAGCGCCTATATGTTTCC
GAATCTCACTTACGTTCAGTCTTTAAAAAATACTCAAATGTTTCCTTACAACATTATATTCTAAGTACAAAAATC
AAAGAAGCTCAACTACTCTTAAAACGAGGAATTCCTGTTGGAGAAGTGGCTAAAAGCTTATATTTTTATGACACT
ACCCATTTTCATAAAATCTTTAAAAAATACACGGGTATTTCTTCAAAAGACTATCTTGCTAAATACCGAGATAAT
ATTTAA

SEQ ID NO: 119

MVIFDLKHVQTLHSLSQLPISVMSQDKALIQVYGNDDYLLCYYQFLKHLAIPQAAQDVIFYEGLFEESFMIFPLC
HYIIAIGPFYPYSLNKDYQEQLANNFLKHSSHRSKEELLSYMALVPHFPINNVRNLLIAIDAFFDTQFETTCQQT

IHQLLQHSKQMTADEDTHRIKHTSKASSQHPPVLEHLNHIMDLVKLGNPQLLKQEINRIPLSSITSSSISALRA
EKNLTVIYLTRLLEFSFVENTDVAKHYSLVKYYMALNEEASDLLKVLRIRCAAIIHFSESLTNKSISDKRQMYNS
VLHYVDSHLYSKLKVSDIAKRLYVSESHLRSVFKKYSNVSLQHYILSTKIKEAQLLLKRGIPVGEVAKSLYFYDT
THFHKIFKKYTGISSKDYLAKYRDNI

19224141 is thought to be a protein F2 fibronectin binding protein. An example of a nucleotide sequence encoding the protein F2 fibronectin binding protein (SEQ ID NO: 120) and a protein F2 fibronectin binding protein amino acid sequence (SEQ ID NO: 121) are set forth below.

SEQ ID NO: 120

ATGACACAAAAAAATAGCTATAAGTTAAGCTTCCTGTTATCCCTAACAGGATTTATTTTAGGTTTATTATTGGTT
TTTATAGGATTGTCCGGAGTATCAGTAGGACATGCGGAAACAAGAAATGGAGCAAACAAACAAGGATCTTTTGAA
ATCAAGAAAGTCGACCAAAACAATAAGCCTTTACCGGGAGCAACTTTTTCACTGACATCAAAGGATGGCAAGGGA
ACATCTGTTCAAACGTTCACTTCAAATGATAAAGGTATTGTAGATGCTCAAAATCTCCAACCAGGGACTTATACC
TTAAAAGAAGAAACAGCACCAGATGGTTATGATAAAACCAGCCGGACTTGGACAGTGACTGTTTATGAGAACGGC
TATACCAAGTTGGTTGAAAATCCCTATAATGGGGAAATCATCAGTAAAGCAGGGTCAAAAGATGTTAGTAGTTCT
TTACAGTTGGAAAATCCCAAAATGTCAGTTGTTTCTAAATATGGGAAAACAGAGGTTAGTAGTGGCGCAGCGGAT
TTCTACCGCAACCATGCCGCCTATTTTAAAATGTCTTTTGAGTTGAAACAAAAGGATAAATCTGAAACAATCAAC
CCAGGTGATACCTTTGTGTTACAGCTGGATAGACGTCTCAATCCTAAAGGTATCAGTCAAGATATCCCTAAAATC
ATTTACGACAGTGCAAATAGTCCGCTTGCGATTGGAAAATACCATGCTGAGAACCATCAACTTATCTATACTTTC
ACAGATTATATTGCGGGTTTAGATAAAGTCCAGTTGTCTGCAGAATTGAGCTTATTCCTAGAGAATAAGGAAGTG
TTGGAAAATACTAGTATCTCAAATTTTAAGAGTACCATAGGTGGGCAGGAGATCACCTATAAAGGAACGGTTAAT
GTTCTTTATGGAAATGAGAGCACTAAAGAAAGCAATTATATTACTAATGGATTGAGCAATGTGGGTGGGAGTATT
GAAAGCTACAACACCGAAACGGGAGAATTTGTCTGGTATGTTTATGTCAATCCAAACCGTACCAATATTCCTTAT
GCGACCATGAATTTATGGGGATTTGGAAGGGCTCGTTCAAATACAAGCGACTTAGAAAACGACGCTAATACAAGT
AGTGCTGAGCTTGGAGAGATTCAGGTCTATGAAGTACCTGAAGGAGAAAAATTACCATCAAGTTATGGGGTTGAT
GTTACAAAACTTACTTTAAGAACGGATATCACAGCAGGCCTAGGAAATGGTTTTCAAATGACCAAACGTCAGCGA
ATTGACTTTGGAAATAATATCCAAAATAAAGCATTTATCATCAAAGTAACAGGGAAAACAGACCAATCTGGTAAG
CCATTGGTTGTTCAATCCAATTTGGCAAGTTTTCGTGGTGCTTCTGAATATGCTGCTTTTACTCCAGTTGGAGGA
AATGTCTACTTCCAAAACGAAATTGCCTTGTCTCCTTCTAAGGGTAGTGGTTCTGGGAAAAGTGAATTTACTAAG
CCCTCTATTACAGTAGCAAATCTAAAACGAGTGGCTCAGCTTCGCTTTAAGAAAATGTCAACTGACAATGTGCCA
TTGCCAGAAGCGGCTTTTGAGCTGCGTTCATCAAATGGTAATAGTCAGAAATTAGAAGCCAGTTCAAACACACAA
GGAGAGGTTCACTTTAAGGACCTGACCTCGGGCACATATGACCTGTATGAAACAAAAGCGCCAAAAGGTTATCAG
CAGGTGACAGAGAAATTGGCGACCGTTACTGTTGATACTACCAAACCTGCTGAGGAAATGGTCACTTGGGGAAGC
CCACATTCGTCTGTAAAAGTAGAAGCTAACAAAGAAGTCACGATTGTCAACCATAAAGAAACCCTTACGTTTTCA
GGGAAGAAAATTTGGGAGAATGACAGACCAGATCAACGCCCAGCAAAGATTCAAGTGCAACTGTTGCAAAATGGT
CAAAAGATGCCTAACCAGATTCAAGAAGTAACGAAGGATAACGATTGGTCTTATCACTTCAAAGACTTGCCTAAG
TACGATGCCAAGAATCAGGAGTATAAGTACTCAGTTGAAGAAGTAAATGTTCCAGACGGCTACAAGGTGTCGTAT
TTAGGAAATGATATATTTAACACCAGAGAAACAGAATTTGTGTTTGAACAGAATAACTTTAACCTTGAATTTGGA
AATGCTGAAATAAAAGGTCAATCTGGGTCAAAAATCATTGATGAAGACACGCTAACGTCTTTCAAAGGTAAGAAA
ATTTGGAAAAATGATACGGCAGAAAATCGTCCCCAAGCCATTCAAGTGCAGCTTTATGCTGATGGAGTGGCTGTG
GAAGGTCAAACCAAATTTATTTCTGGCTCAGGTAATGAGTGGTCATTTGAGTTTAAAAACTTGAAGAAGTATAAT
GGAACAGGTAATGACATCATTTACTCAGTTAAAGAAGTAACTGTTCCAACAGGTTATGATGTGACTTACTCAGCT
AATGATATTATTAATACCAAACGTGAGGTTATTACACAACAAGGACCGAAACTAGAGATTGAAGAAACGCTTCCG
CTAGAATCAGGTGCTTCAGGCGGTACCACTACTGTCGAAGACTCACGCCCAGTTGATACCTTATCAGGTTTATCA
AGTGAGCAAGGTCAGTCCGGTGATATGACAATTGAAGAAGATAGTGCTACCCATATTAAATTCTCAAAACGTGAT
ATTGACGGCAAAGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTCATCTGGTAAAACTATTAGTACATGGATT
TCAGATGGACAAGTGAAAGATTTCTACCTGATGCCAGGAAAATATACATTTGTCGAAACCGCAGCACCAGACGGT
TATGAGATAGCAACTGCTATTACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAA
GGTGACACTCATATTGTCATGGTTGATGCTTACAAGCCAACTAAGGGTTCAGGTCAGGTTATTGATATTGAAGAA
AAGCTTCCAGACGAGCAAGGTCATTCTGGTTCAACTACTGAAATAGAAGACAGTAAATCTTCAGACCTTATCATT
GGCGGTCAAGGTGAAGTTGTTGACACAACAGAAGACACACAAAGTGGTATGACGGGCCATTCTGGCTCAACTACT
GAAATAGAAGATAGCAAGTCTTCAGACGTTATCATTGGTGGTCAGGGGCAGGTTGTCGAGACAACAGAGGATACC
CAAACTGGCATGTACGGGGATTCTGGTTGTAAAACGGAAGTCGAAGATACTAAACTAGTACAATCCTTCCACTTT
GATAACAAGGAACCAGAAAGTAACTCTGAGATTCCTAAAAAAGATAAGCCAAAGAGTAATACTAGTTTACCAGCA
ACTGGTGAGAAGCAACATAATATGTTCTTTTGGATGGTTACTTCTTGCTCACTTATTAGTAGTGTTTTTGTAATA
TCACTAAAATCCAAAAAACGCCTATCATCATGTTAA

SEQ ID NO: 121

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKG
TSVQTFTSNDKGIVDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIISKAGSKDVSSS
LQLENPKMSVVSKYGKTEVSSGAADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLDRRLNPKGISQDIPKI
IYDSANSPLAIGKYHAENHQLIYTFTDYIAGLDKVQLSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTVN
VLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPYATMNLWGFGRARSNTSDLENDANTS
SAELGEIQVYEVPEGEKLPSSYGVDVTKLTLRTDITAGLGNGFQMTKRQRIDFGNNIQNKAFIIKVTGKTDQSGK
PLVVQSNLASFRGASEYAAFTPVGGNVYFQNEIALSPSKGSGSGKSEFTKPSITVANLKRVAQLRFKKMSTDNVP
LPEAAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLYETKAPKGYQQVTEKLATVTVDTTKPAEEMVTWGS
PHSSVKVEANKEVTIVNHKETLTFSGKKIWENDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPK
YDAKNQEYKYSVEEVNVPDGYKVSYLGNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKK
IWKNDTAENRPQAIQVQLYADGVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSA
NDIINTKREVITQQGPKLEIEETLPLESGASGGTTTVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRD
IDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEIATAITFTVNEQGQVTVNGKATK
GDTHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDLIIGGQGEVVDTTEDTQSGMTGHSGSTT
EIEDSKSSDVIIGGQGQVVETTEDTQTGMYGDSGCKTEVEDTKLVQSFHFDNKEPESNSEIPKKDKPKSNTS*LPA*
*TG*EKQHNMFFWMVTSCSLISSVFVISLKSKKRLSSC

19224141 contains an amino acid motif indicative of a cell wall anchor: **SEQ ID NO: 181** LPATG (shown in italics in SEQ ID NO: 121, above). In some recombinant host cell systems, it may be preferable to remove this motif to facilitate secretion of a recombinant 19224141 protein from the host cell. Alternatively, in other recombinant host cell systems, it may be preferable to use the cell wall anchor motif to anchor the recombinantly expressed protein to the cell wall. The extracellular domain of the expressed protein may be cleaved during purification or the recombinant protein may be left attached to either inactivated host cells or cell membranes in the final composition.

Two pilin motifs, discussed above, containing conserved lysine (K) residues have also been identified in 19224141. The pilin motif sequences are underlined in SEQ ID NO: 121, below. Conserved lysine (K) residues are also marked in bold, at amino acid residues 157 and 163 and at amino acid residues 216, 224, and 238. The pilin sequence, in particular the conserved lysine residues, are thought to be important for the formation of oligomeric, pilus-like structures. Preferred fragments of 19224141 include at least one conserved lysine residue. Preferably, fragments include at least one pilin sequence.

SEQ ID NO: 121

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKG
TSVQTFTSNDKGIVDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIISKAGSKDVSSS
LQLENPKMSVVSKYGKTEVSSGAADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLDRRLNPKGISQDIPKI
IYDSANSPLAIGKYHAENHQLIYTFTDYIAGLDKVQLSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTVN
VLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPYATMNLWGFGRARSNTSDLENDANTS
SAELGEIQVYEVPEGEKLPSSYGVDVTKLTLRTDITAGLGNGFQMTKRQRIDFGNNIQNKAFIIKVTGKTDQSGK
PLVVQSNLASFRGASEYAAFTPVGGNVYFQNEIALSPSKGSGSGKSEFTKPSITVANLKRVAQLRFKKMSTDNVP
LPEAAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLYETKAPKGYQQVTEKLATVTVDTTKPAEEMVTWGS
PHSSVKVEANKEVTIVNHKETLTFSGKKIWENDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPK
YDAKNQEYKYSVEEVNVPDGYKVSYLGNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKK
IWKNDTAENRPQAIQVQLYADGVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSA
NDIINTKREVITQQGPKLEIEETLPLESGASGGTTTVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRD
IDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEIATAITFTVNEQGQVTVNGKATK
GDTHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDLIIGGQGEVVDTTEDTQSGMTGHSGSTT
EIEDSKSSDVIIGGQGQVVETTEDTQTGMYGDSGCKTEVEDTKLVQSFHFDNKEPESNSEIPKKDKPKSNTSLPA
TGEKQHNMFFWMVTSCSLISSVFVISLKSKKRLSSC

Two E boxes containing conserved glutamic residues have been identified in 19224141. The E-box motifs are underlined in SEQ ID NO: 121, below. The conserved glutamic acid (E) residues, at

amino acid residues 367 and 944, are marked in bold. The E box motifs, in particular the conserved glutamic acid residues, are thought to be important for the formation of oligomeric pilus-like structures of 19224141. Preferred fragments of 19224141 include at least one conserved glutamic acid residue. Preferably, fragments include at least one E box motif.

SEQ ID NO: 121

MTQKNSYKLSFLLSLTGFILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKG
TSVQTFTSNDKGIVDAQNLQPGTYTLKEETAPDGYDKTSRTWTVTVYENGYTKLVENPYNGEIISKAGSKDVSSS
LQLENPKMSVVSKYGKTEVSSGAADFYRNHAAYFKMSFELKQKDKSETINPGDTFVLQLDRRLNPKGISQDIPKI
IYDSANSPLAIGKYHAENHQLIYTFTDYIAGLDKVQLSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTVN
VLYGNESTKESNYITNGLSNVGGSIESYNTETGEFVWYVYVNPNRTNIPYATMNLWGFGRARSNTSDLENDANTS
SAELGEIQVYEVPEGEKLPSSYGVDVTKLTLRTDITAGLGNGFQMTKRQRIDFGNNIQNKAFIIKVTGKTDQSGK
PLVVQSNLASFRGASEYAAFTPVGGNVYFQNEIALSPSKGSGSGKSEFTKPSITVANLKRVAQLRFKKMSTDNVP
LPEAAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLYETKAPKGYQQVTEKLATVTVDTTKPAEEMVTWGS
PHSSVKVEANKEVTIVNHKETLTFSGKKIWENDRPDQRPAKIQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPK
YDAKNQEYKYSVEEVNVPDGYKVSYLGNDIFNTRETEFVFEQNNFNLEFGNAEIKGQSGSKIIDEDTLTSFKGKK
IWKNDTAENRPQAIQVQLYADGVAVEGQTKFISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSA
NDIINTKREVITQQGPKLEIEETLPLESGASGGTTTVEDSRPVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRD
IDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEIATAITFTVNEQGQVTVNGKATK
GDTHIVMVDAYKPTKGSGQVIDIEEKLPDEQGHSGSTTEIEDSKSSDLIIGGQGEVVDTTEDTQSGMTGHSGSTT
EIEDSKSSDVIIGGQGQVVETTEDTQTGMYGDSGCKTEVEDTKLVQSFHFDNKEPESNSEIPKKDKPKSN
TSLPATGEKQHNMFFWMVTSCSLISSVFVISLKSKKRLSSC

As discussed above, applicants have also determined the nucleotide and encoded amino acid sequence of fimbrial structural subunits in several other GAS AI-4 strains of bacteria. Examples of sequences of these fimbrial structural subunits are set forth below.

M12 strain isolate 20010296 is a GAS AI-4 strain of bacteria. 20010296_fimbrial is thought to be a fimbrial structural subunit of M12 strain isolate 20010296. An example of a nucleotide sequence encoding the 20010296_fimbrial protein (SEQ ID NO: 257) and a 20010296_fimbrial protein amino acid sequence (SEQ ID NO: 258) are set forth below.

SEQ ID NO: 257

agcagtggtcaattaacaataaaaaaatcaattacaaattttaatgatgatacacttttg
atgcctaagacagactatacttttagcgttaatccggatagtgcggctacaggtactgaa
agtaatttaccaattaaaccaggtattgctgttaacaatcaagatattaaggtttcttat
tctaatactgataagacatcaggtaaagaaaaacaagttgttgttgactttatgaaagtt
acttttcctagcgttggtatttaccgttatgttgttaccgagaataaagggacagcagaa
ggagttacatatgatgatacaaaatggttagttgacgtctatgttggtaataatgaaaag
ggaggtcttgaaccaaagtatattgtatctaaaaaaggagattctgctactaaagaacca
atccagtttaataattcattcgaaacaacgtcattaaaaattgaaaaggaagttactggt
aatacaggagatcataaaaaagcatttaactttacattaacattgcaaccaaatgaatac
tatgaggcaagttcggttgtgaaaattgaagagaacggacaaacgaaagatgtgaaaatt
ggggaggcatataagtttactttgaacgatagtcagagtgtgatattgtctaaattacca
gttggtattaattataaagttgaagaagcagaagctaatcaaggtggatatactacaaca
gcaactttaaaagatggagaaaagttatctacttataacttaggtcaggaacataaaaca
gacaagactgctgatgaaatcgt

SEQ ID NO: 258

SSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATGTESNLP
IKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENKGTAEGV
TYDDTKWLVDVYVGNNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKIEKEVTG
NTGDHKKAFNFTLTLQPNEYYEASSVVKIEENGQTKDVKIGEAYKFTLNDSQSVILSK
LPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADEIV

PCT M12 strain isolate 20020069 is a GAS AI-4 strain of bacteria. 20020069_fimbrial is thought to be a fimbrial structural subunit of M12 strain isolate 20020069. An example of a nucleotide sequence encoding the 20020069_fimbrial protein (SEQ ID NO: 259) and a 20020069_fimbrial protein amino acid sequence (SEQ ID NO: 260) are set forth below.

SEQ ID NO: 259

agcagtggtcaattaacaataaaaaaatcaattacaaattttaatgatgatacacttttg
atgcctaagacagactatacttttagcgttaatccggatagtgcggctacaggtactgaa
agtaatttaccaattaaaccaggtattgctgttaacaatcaagatattaaggtttcttat
tctaatactgataagacatcaggtaaagaaaaacaagttgttgttgactttatgaaagtt
actttcctagcgttggtatttaccgttatgttgttaccgagaataaagggacagcagaa
ggagttacatatgatgatacaaaatggttagttgacgtctatgttggtaataatgaaaag
ggaggtcttgaaccaaagtatattgtatctaaaaaaggagattctgctactaaagaacca
atccagtttaataattcattcgaaacaacgtcattaaaaattgaaaaggaagttactggt
aatacaggagatcataaaaaagcatttaactttacattaacattgcaaccaaatgaatac
tatgaggcaagttcggttgtgaaaattgaagagaacggacaaacgaaagatgtgaaaatt
ggggaggcatataagtttactttgaacgatagtcagagtgtgatattgtctaaattacca
gttggtattaattataaagttgaagaagcagaagctaatcaaggtggatatactacaaca
gcaactttaaaagatggagaaaagttatctacttataacttaggtcaggaacataaaaca
gacaagactgctgatgaaatcgt

SEQ ID NO: 260

SSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATGTESNLP
IKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENKGTAEGV
TYDDTKWLVDVYVGNNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKIEKEVTG
NTGDHKKAFNFTLTLQPNEYYEASSVVKIEENGQTKDVKIGEAYKFTLNDSQSVILSK
LPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADEIV

M12 strain isolate CDC SS 635 is a GAS AI-4 strain of bacteria. CDC SS 635_fimbrial is thought to be a fimbrial structural subunit of M12 strain isolate CDC SS 635. An example of a nucleotide sequence encoding the CDC SS 635_fimbrial protein (SEQ ID NO: 261) and a CDC SS 635_fimbrial protein amino acid sequence (SEQ ID NO: 262) are set forth below.

SEQ ID NO: 261

gagacggcaggggttgttagcagtggtcaattaacaataaaaaaatcaattacaaatttt
aatgatgatacacttttgatgcctaagacagactatacttttagcgttaatccggatagt
gcggctacaggtactgaaagtaatttaccaattaaaccaggtattgctgttaacaatcaa
gatattaaggtttcttattctaatactgataagacatcaggtaaagaaaaacaagttgtt
gttgactttatgaaagttacttttcctagcgttggtatttaccgttatgttgttaccgag
aataaagggacagcagaaggagttacatatgatgatacaaaatggttagttgacgtctat
gttggtaataatgaaaagggaggtcttgaaccaaagtatattgtatctaaaaaaggagat
tctgctactaaagaaccaatccagtttaataattcattcgaaacaacgtcattaaaaatt
gaaaaggaagttactggtaatacaggagatcataaaaaagcatttaactttacattaaca
ttgcaaccaaatgaatactatgaggcaagttcggttgtgaaaattgaagagaacggacaa
acgaaagatgtgaaaattggggaggcatataagtttactttgaacgatagtcagagtgtg
atattgtctaaattaccagttggtattaattataaagttgaagaagcagaagctaatcaa
ggtggatatactacaacagcaactttaaaagatggagaaaagttatctacttataactta
ggtcaggaacataaaacagacaagactgctgatgaaatcgttgtcacaaataaccgtgac
act

SEQ ID NO: 262

ETAGVVSSGQLTIKKSITNFNDDTLLMPKTDYTFSVNPDSAATG
TESNLPIKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIYRYVVTENK
GTAEGVTYDDTKWLVDVYVGNNEKGGLEPKYIVSKKGDSATKEPIQFNNSFETTSLKI
EKEVTGNTGDHKKAFNFTLTLQPNEYYEASSVVKIEENGQTKDVKIGEAYKFTLNDSQ
SVILSKLPVGINYKVEEAEANQGGYTTTATLKDGEKLSTYNLGQEHKTDKTADEIVVT
NNRDT

PCT/US2005/027239

M5 strain isolate ISS4883 is a GAS AI-4 strain of bacteria. ISS4883_fimbrial is thought to be a fimbrial structural subunit of M5 strain isolate ISS 4883. An example of a nucleotide sequence encoding the ISS4883_fimbrial protein (SEQ ID NO: 265) and an ISS4883_fimbrial protein amino acid sequence (SEQ ID NO: 266) are set forth below.

SEQ ID NO: 265

```
gagacggcaggggttgtaacaggaaaatcactacaagttacaaagacaatgacttatgat
gatgaagaggtgttaatgcccgaaaccgcctttacttttactatagagcctgatatgact
gcaagtggaaaagaaggcgacctagatattaaaaatggaattgtagaaggcttagacaaa
caagtaacagtaaaatataagaatacagataaaacatctcaaaaaactaaaatagcacaa
tttgatttttctaaggttaaatttccagctataggtgtttaccgctatatggtttcagag
aaaaacgataaaaaagacggaattaggtacgatgataaaaagtggactgtagatgtttat
gttgggaataaggccaataacgaagaaggtttcgaagttctatatattgtatcaaaagaa
ggtacttctagtactaaaaaaccaattgaatttacaaactctattaaaactacttcctta
aaaattgaaaaacaaataactggcaatgcaggagatcgtaaaaaatcattcaacttcaca
ttaacattacaaccaagtgaatattataaaaccggatcagttgtgaaaatcgaacaggat
ggaagtaaaaaagatgtgacgataggaacgccttacaaatttacttttgggacacggtaag
agtgtcatgttatcgaaattaccaattggtatcaattactatcttagtgaagacgaagcg
aataaagacggttacactacaacggcaacattaaaagaacaaggcaaagaaaagagttcc
gatttcactttgagtactcaaaaccagaaaacagacgaatctgctgacgaaatcgttgtc
acaaataagcgtgacactctcgag
```

SEQ ID NO: 266

```
ETAGVVTGKSLQVTKTMTYDDEEVLMPETAFTFTIEPDMTASGK
EGDLDIKNGIVEGLDKQVTVKYKNTDKTSQKTKIAQFDFSKVKFPAIGVYRYMVSEKN
DKKDGIRYDDKKWTVDVYVGNKANNEEGFEVLYIVSKEGTSSTKKPIEFTNSIKTTSL
KIEKQITGNAGDRKKSFNFTLTLQPSEYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGH
GKSVMLSKLPIGINYYLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQNQKTDESAD
EIVVTNKRDTLE
```

M50 strain isolate ISS4538 is a GAS AI-4 strain of bacteria. ISS4538_fimbrial is thought to be a fimbrial structural subunit of M50 strain ISS 4538. An example of a nucleotide sequence encoding the ISS4538_fimbrial protein (SEQ ID NO: 255) and an ISS4538_fimbrial protein amino acid sequence (SEQ ID NO: 256) are set forth below.

SEQ ID NO: 255

```
atgaaaaaaaataaattattacttgctactgcaatcttagcaactgctttaggaacagct
tctttaaatcaaaacgtaaaagctgagacggcaggggttgttagcagtggtcaattaaca
ataaaaaaatcaattacaaattttaatgatgatacacttttgatgcctaagacagactat
acttttagcgttaatccggatagtgcggctacaggtactgaaagtaatttaccaattaaa
ccaggtattgctgttaacaatcaagatattaaggtttcttattctaatactgataagaca
tcaggtaaagaaaaacaagttgttgttgacttttatgaaagttacttttcctagcgttggt
atttaccgttatgttgttaccgagaataaagggacagcagaaggagttacatatgatgat
acaaaatggttagttgacgtctatgttggtaataatgaaaagggaggtcttgaaccaaag
tatattgtatctaaaaaaggagattctgctactaaagaaccaatccagtttaataattca
ttcgaaacaacgtcattaaaaattgaaaagaaagttactggtaatacaggagatcataaa
aaagcatttaactttacattaacattgcaaccaaatgaatactatgaggcaagttcggtt
gtgaaaattgaagagaacggacaaacgaaagatgtgaaaattggggaggcatataagttt
actttgaacgatagtcagagtgtgatattgtctaaattaccagttggtattaattataaa
gttgaagaagcagaagctaatcaaggtggatatactacaacagcaactttaaaagatgga
gaaaagttatctacttataacttaggtcaggaacataaaacagacaagactgctgatgaa
atcgttgtcacaaataancgngacactcnagttccaacnggtgtngtaggcaccccnnct
ccattcncagttcttancattgnggctantggtggngtnatntatnttacaaaacgnaaa
aaagnataa
```

SEQ ID NO: 256

```
MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKS
ITNFNDDTLLMPKTDYTFSVNPDSAATGTESNLPIKPGIAVNNQDIKVSYSNTDKTSG
```

KPKQVVVDFMKVTPPSVGLYRVVVTENKGTAEGVTYDDTKWLVDVYVGNNEKGGLEPK
YIVSKKGDSATKEPIQFNNSFETTSLKIEKKVTGNTGDHKKAFNFTLTLQPNEYYEAS
SVVKIEENGQTKDVKIGEAYKFTLNDSQSVILSKLPVGINYKVEEAEANQGGYTTTAT
LKDGEKLSTYNLGQEHKTDKTADEIVVTNXRDTXVPTGVVGTPPPFXVLXIXAXGGVX
YXTKRKKX

There may be an upper limit to the number of GAS proteins which will be in the compositions of the invention. Preferably, the number of GAS proteins in a composition of the invention is less than 20, less than 19, less than 18, less than 17, less than 16, less than 15, less than 14, less than 13, less than 12, less than 11, less than 10, less than 9, less than 8, less than 7, less than 6, less than 5, less than 4, or less than 3. Still more preferably, the number of GAS proteins in a composition of the invention is less than 6, less than 5, or less than 4. Still more preferably, the number of GAS proteins in a composition of the invention is 3.

The GAS proteins and polynucleotides used in the invention are preferably isolated, *i.e.*, separate and discrete, from the whole organism with which the molecule is found in nature or, when the polynucleotide or polypeptide is not found in nature, is sufficiently free of other biological macromolecules so that the polynucleotide or polypeptide can be used for its intended purpose.

Examples Other Gram positive bacterial Adhesin Island Sequences

The Gram positive bacteria AI polypeptides of the invention can, of course, be prepared by various means (*e.g.* recombinant expression, purification from a gram positive bacteria, chemical synthesis *etc.*) and in various forms (*e.g.* native, fusions, glycosylated, non-glycosylated *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other streptococcal or host cell proteins) or substantially isolated form.

The Gram positive bacteria AI proteins of the invention may include polypeptide sequences having sequence identity to the identified Gram positive bacteria proteins. The degree of sequence identity may vary depending on the amino acid sequence (a) in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more). Polypeptides having sequence identity include homologs, orthologs, allelic variants and mutants of the identified Gram positive bacteria proteins. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affinity gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The Gram positive bacteria adhesin island polynucleotide sequences may include polynucleotide sequences having sequence identity to the identified Gram positive bacteria adhesin island polynucleotide sequences. The degree of sequence identity may vary depending on the polynucleotide sequence in question, but is preferably greater than 50% (*e.g.* 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more).

The Gram positive bacteria adhesin island polynucleotide sequences of the invention may include polynucleotide fragments of the identified adhesin island sequences. The length of the fragment may vary depending on the polynucleotide sequence of the specific adhesin island sequence, but the fragment is preferably at least 10 consecutive polynucleotides, (*e.g.* at least 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more).

The Gram positive bacteria adhesin island amino acid sequences of the invention may include polypeptide fragments of the identified Gram positive bacteria proteins. The length of the fragment may vary depending on the amino acid sequence of the specific Gram positive bacteria antigen, but the fragment is preferably at least 7 consecutive amino acids, (*e.g.* 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). Preferably the fragment comprises one or more epitopes from the sequence. The fragment may comprise at least one T-cell or, preferably, a B-cell epitope of the sequence. T- and B-cell epitopes can be identified empirically (e.g., using PEPSCAN [Geysen *et al.* (1984) *PNAS USA* 81:3998-4002; Carter (1994) *Methods Mol. Biol.* 36:207-223, or similar methods], or they can be predicted (*e.g.*, using the Jameson-Wolf antigenic index [Jameson, BA *et al.* 1988, *CABIOS* 4(1):1818-186], matrix-based approaches [Raddrizzani and Hammer (2000) *Brief Bioinform.* 1(2):179-189], TEPITOPE [De Lalla *et al.* (199) *J. Immunol.* 163:1725-1729], neural networks [Brusic *et al.* (1998) *Bioinformatics* 14(2):121-130], OptiMer & EpiMer [Meister *et al.* (1995) *Vaccine* 13(6):581-591; Roberts *et al.* (1996) *AIDS Res. Hum. Retroviruses* 12(7):593-610], ADEPT [Maksyutov & Zagrebelnaya (1993) *Comput. Appl. Biosci.* 9(3):291-297], Tsites [Feller & de la Cruz (1991) *Nature* 349(6311):720-721], hydrophilicity [Hopp (1993) *Peptide Research* 6:183-190], antigenic index [Welling *et al.* (1985)*FEBS Lett.* 188:215-218] or the methods disclosed in Davenport *et al.* (1995) *Immunogenetics* 42:392-297, etc. Other preferred fragments include (1) the N-terminal signal peptides of each identified Gram positive bacteria protein, (2) the identified Gram positive bacteria protein without their N-terminal signal peptides, (3) each identified Gram positive bacteria protein wherein up to 10 amino acid residues (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) are deleted from the N-terminus and/or the C-terminus *e.g.* the N-terminal amino acid residue may be deleted. Other fragments omit one or more domains of the protein (*e.g.* omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain), and (4) the polypeptides, but without their N-terminal amino acid residue.

As indicated in the above text, nucleic acids and polypeptides of the invention may include sequences that:

(a) are identical (*i.e.*, 100% identical) to the sequences disclosed in the sequence listing;

(b) share sequence identity with the sequences disclosed in the sequence listing;

(c) have 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 single nucleotide or amino acid alterations (deletions, insertions, substitutions), which may be at separate locations or may be contiguous, as compared to the sequences of (a) or (b);

(d) when aligned with a particular sequence from the sequence listing using a pairwise alignment algorithm, a moving window of x monomers (amino acids or nucleotides)

moving from start (N-terminus or 5′) to end (C-terminus or 3′), such that for an alignment that extends to p monomers (where $p>x$) there are $p-x+1$ such windows, each window has at least $x \cdot y$ identical aligned monomers, where: x is slected from 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200; y is selected from 0.50, 0.60, 0.70, 0.75, 0.80, 0.85, 0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99; and if $x \cdot y$ is not an integer then it is rounded up to the nearest integer. The preferred pairwise alignment algorithm is the Needleman-Wunsch global alignment algorithm [Needlman &Wunsch (1970) *J. Mol. Biol.* 48, 443-453], using default parameters (*e.g.*, with Gap opening penalty = 10.0, and with Gap extension penalty = 0.5, using the EBLOSUM62 scoring matrix). This algorithm is conveniently implemented in the *needle* tool in the EMBOSS package [Rice *et al.* (2000) *Trends Genet.* 16:276-277].

The nucleic acids and polypeptides of the inention may additionally have further sequences to the N-terminus/5′ and/or C-terminus/3′ of these sequences (a) to (d).

All of the Gram positive bacterial sequences referenced herein are publicly available through PubMed on GenBank.

Streptococcus pneumoniae Adhesin Island Sequences

As discussed above, a *S. pneumoniae* AI sequence is present in the TIGR4 *S. pneumoniae* genome. Examples of *S. pneumoniae* AI sequences are set forth below.

SrtD (Sp0468) is a sortase. An example of an amino acid sequence of SrtD is set forth in SEQ ID NO: 80.

SEQ ID NO: 80

MSRTKLRALLGYLLMLVACLIPIYCFGQMVLQSLGQVKGHATFVKSMTTEMYQEQQNHSLAYNQRLASQNRIVDP
FLAEGYEVNYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLGMGLAHVDGTPLPLDGTGIRSVIAGHRAEPSH
VFFRHLDQLKVGDALYYDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAV
YQKSDPQTAAVARVAFTKEGQSVSRVATSQWLYRGLVVLAFLGILFVLWKLARLLRGK

SrtC (Sp0467) is a sortase. An example of an amino acid sequence of SrtC is set forth in SEQ ID NO: 81.

SEQ ID NO: 81

MSRYYYRIESNEVIKEFDETVSQMDKAELEERWRLAQAFNATLKPSEILDPFTEQEKKKGVSEYANMLKVHERIG
YVEIPAIDQEIPMYVGTSEDILQKGAGLLEGASLPVGGENTHTVITAHRGLPTAELFSQLDKMKKGDIFYLHVLD
QVLAYQVDQIVTVEPNDFEPVLIQHGEDYATLLTCTPYMINSHRLLVRGKRIPYTAPIAERNRAVRERGQFWLWL
LLGAMAVILLLLYRVYRNRRIVKGLEKQLEGRHVKD

SrtB (SP0466) is a sortase. An example of an amino acid sequence of SrtB is set forth in SEQ ID NO: 82.

SEQ ID NO: 82

MAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDSLNNVVSGDPWSEEMKKKGRAEYARM
LEIHERMGHVEIPVIDVDLPVYAGTAEEVLQQGAGHLEGTSLPIGGNSTHAVITAHTGLPTAKMFTDLTKLKVGD
KFYVHNIKEVMAYQVDQVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINTHRLLVRGHRIPYVAEVEEEFIAANK
LSHLYRYLFYVAVGLIVILLWIIRRLRKKKKQPEKALKALKAARKEVKVEDGQQ

Sp0465 is a hypothetical protein. An example of an amino acid sequence of Sp0465 is set forth in SEQ ID NO: 83.

PCT/US05/27239

SEQ ID NO: 83

MFLPFLSASLYLQTHHFIAFPNRQSYLLRETRKSHFFLIHHPF

RrgC (SP0464) is a cell wall surface anchor family protein. RrgC contains a sortase substrate motif VPXTG (SEQ ID NO: 137), shown in italics in SEQ ID NO: 84.

SEQ ID NO: 84

MISRIFFVMALCFSLVWGAHAVQAQEDHTLVLQLENYQEVVSQLPSRDGHRLQVWKLDDSYSYDDRVQIVRDLHS
WDENKLSSFKKTSFEMTFLENQIEVSHIPNGLYYVRSIIQTDAVSYPAEFLFEMTDQTVEPLVIVAKKTDTMTTK
VKLIKVDQDHNRLEGVGFKLVSVARDVSEKEVPLIGEYRYSSSGQVGRTLYTDKNGEIFVTNLPLGNYRFKEVEP
LAGYAVTTLDTDVQLVDHQLVTITVVNQKLPRGNVDFMKVDGRTNTSLQGAMFKVMKEESGHYTPVLQNGKEVVV
TSGKDGRFRVEGLEYGTYYLWELQAPTGYVQLTSPVSFTIGKDTRKELVTVVKNNKRPRID*VPDTG*EETLYILML
VAILLFGSGYYLTKKPNN

RrgB (Sp0463) is a cell wall surface anchor protein. RrgB contains a sortase substrate motif IPXTG (SEQ ID NO: 133), shown in italics in SEQ ID NO: 85.

SEQ ID NO: 85

MKSINKFLTMLAALLLTASSLFSAATVFAAGTTTTSVTVHKLLATDGDMDKIANELETGNYAGNKVGVLPANAKE
IAGVMFVWTNTNNEIIDENGQTLGVNIDPQTFKLSGAMPATAMKKLTEAEGAKFNTANLPAAKYKIYEIHSLSTY
VGEDGATLTGSKAVPIEIELPLNDVVDAHVYPKNTEAKPKIDKDFKGKANPDTPRVDKDTPVNHQVGDVVEYEIV
TKIPALANYATANWSDRMTEGLAFNKGTVKVTVDDVALEAGDYALTEVATGFDLKLTDAGLAKVNDQNAEKTVKI
TYSATLNDKAIVEVPESNDVTFNYGNNPDHGNTPKPNKPNENGDLTLTKTWVDATGAPIPAGAEATFDLVNAQTG
KVVQTVTLTTDKNTVTVNGLDKNTEYKFVERSIKGYSADYQEITTAGEIAVKNWKDENPKPLDPTEPKVVTYGKK
FVKVNDKDNRLAGAEFVIANADNAGQYLARKADKVSQEEKQLVVTTKDALDRAVAAYNALTAQQQTQQEKEKVDK
AQAAYNAAVIAANNAFEWVADKDNENVVKLVSDAQGRFEITGLLAGTYYLEETKQPAGYALLTSRQKFEVTATSY
SATGQGIEYTAGSGKDDATKVVNKKIT*IPQTGG*IGTIIFAVAGAAIMGIAVYAYVKNNKDEDQLA

RrgA (Sp0462) is a cell wall surface anchor protein. RrgA contains a sortase substrate motif YPXTG (SEQ ID NO: 186), indicated in italics in SEQ ID NO: 86.

SEQ ID NO: 86

MLNRETHMKKVRKIFQKAVAGLCCISQLTAFSSIVALAETPETSPAIGKVVIKETGEGGALLGDAVFELKNNTDG
TTVSQRTEAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQWTVEVEKNGRTTVQGEQVENREEALSDQYPQTGT
YPDVQTPYQIIKVDGSEKNGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTVYEQKDKSVPL
DVVILLDNSNSMSNIRNKNARRAERAGEATRSLIDKITSDSENRVALVTYASTIFDGTEFTVEKGVADKNGKRLN
DSLFWNYDQTSFTTNTKDYSYLKLTNDKNDIVELKNKVPTEAEDHDGNRLMYQFGATFTQKALMKADEILTQQAR
QNSQKVIFHITDGVPTMSYPINFNHATFAPSYQNQLNAFFSKSPNKDGILLSDFITQATSGEHTIVRGDGQSYQM
FTDKTVYEKGAPAAFPVKPEKYSEMKAAGYAVIGDPINGGYIWLNWRESILAYPFNSNTAKITNHGDPTRWYYNG
NIAPDGYDVFTVGIGINGDPGTDEATATSFMQSISSKPENYTNVTDTTKILEQLNRYFHTIVTEKKSIENGTITD
PMGELIDLQLGTDGRFDPADYTLTANDGSRLENGQAVGGPQNDGGLLKNAKVLYDTTEKRIRVTGLYLGTDEKVT
LTYNVRLNDEFVSNKFYDTNGRTTLHPKEVEQNTVRDFPIPKIRDVRKYPEITISKEKKLGDIEFIKVNKNDKKP
LRGAVFSLQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVN
GEVRDVTSIVPQDIPAGYEFTNDKHYITNEPIPPKRE*YPRTG*GIGMLPFYLIGCMMMGGVLLYTRKHP

RlrA (Sp0461) is a transcriptional regulator. An example of an amino acid sequence for RlrA is set forth in SEQ ID NO: 87.

SEQ ID NO: 87

MLNKYIEKRITDKITILNILLDIRSIELDELSTLTSLQSKSLLSILQELQETFEEELTFNLDTQQVQLIEHHSHQ
TNYYFHQLYNQSTILKILRFFLLQGNQSFNEFTQKEYISIATGYRVRQKCGLLLRSVGLDLVKNQVVGPEYRIRF
LIALLQFHFGIEIYDLNDGSMDWVTHMIVQSNSQLSHELLEITPDEYVHFSILVALTWKRREFPLEFPESKEFEK
LKNLFMYPILMEHCQTYLEPHANMTFTQEELDYIFLVYCSANSSFSKDKWNQEKKTHTIQLILQHTRGKHLLSKF
KNILGNDISNSLSFLTALTFLTRTFLFGLQNLVPYYNYYEHYGIESDKPLYHISKAIVQEWMTEQKIEGVIDQHR
LYLFSLYLTETIFSSLPAIPIFIILNNQADVNLIKSIILRNFTDKVASVTGYNILISPPPSEEHLTEPLIIITTK
EYLPYVKKQYPKGKHHFLTIALDLHVSQQRLIYQTIVDIRKEAFDKRVAMIAKKAHYLL

As discussed above, a *S. pneumoniae* AI sequence is present in the *S. pneumoniae* strain 670 genome. Examples of *S. pneumoniae* AI sequences are set forth below.

Orf1_670 is a transposase. An example of an amino acid sequence of orf1_670 is set forth in SEQ ID NO: 171.

SEQ ID NO: 171

MEHINHTTLLIGIKDKNITLNKAIQHDTHIEVFATLDYHPPKCKHCKGKQIKYDFQKPSKIPFIEIGGFPSLIHL
KKRRFQCKSCRKVTVAETTLVQKNCQISEMVRQKIAQLLLNREALTHIASKLAISTSTSTVYRKLKQFHFQEDYT
TLPEILSWDEFSYQKGKLAFIAQDFNTKKIMTILDNRRQTTIRNHFFKYSKEARKKVKVVTVDMSGSYIPLIKKL
FPNAKIVLDRFHIVQHMSRALNQTRINIMKQFDDKSLEYRALKYYWKFILKDSRKLSLKPFYARTFRETLTPREC
LKKIFTLVPELKDYYDLYQLLLFHLQEKNTDQFWGLIQDTLPHLNRTFKTTLSTFICYKNYITNAIELPYSNAKL
EATNKLIKDIKRNAFGFRNFENFKKRIFIALNIKKERTKFVLSRA

Orf2_670 is a transcriptional regulator. An example of an amino acid sequence of Orf2_670 is set forth in SEQ ID NO: 172.

SEQ ID NO: 172

MLNKYIEKRITDKITILNILLDIRSIELDELSTLTSLQSKSLLSILQELQETFEEELTFNLDTQQVQLIEHHSHQ
TNYYFHQLYNQSTILKILRFFLLQGNQSFNEFTQKEYISIATGYRVRQKCGLLLRSVGLDLVKNQVVGPEYRIRF
LIALLQFHFGIEIYDLNDGSMDWVTHMIVQSNSQLSHELLEITPDEYVHFSILVALTWKRREFPLEFPESKEFEK
LKNLFMYPILMEHCQTYLEPHANMTFTQEELDYIFLVYCSANSSFSKDKWNQEKKTHTIQLILQHTRGKHLLSKF
KNILGNDISNSLSFLTALTFLTRTFLFGLQNLVPYYNYYEHYGIESDKPLYHISKAIVQEWMTEQKIEGVIDQHR
LYLFSLYLTETIFSSLPAIPIFIILNNQADVNLIKSIILRNFTDKVASVTGYNILISPPPSEEHLTEPLIIITTK
EYLPYVKKQYPKGKHHFLTIALDLHVSQQRLIYQTIVDIRKEAFDKRVAMIAKKAHYLL

Orf3_670 is a cell wall surface anchor family proten. An example of an amino acid sequence of Orf3_670 is set forth in SEQ ID NO: 173.

SEQ ID NO: 173

MLNRETHMKKVRKIFQKAVAGLCCISQLTAFSSIVALAETPETSPAIGKVVIKETGEGGALLGDAVFELKNNTDG
TTVSQRTEAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQWTVEVEKNGRTTVQGEQVENREEALSDQYPQTGT
YPDVQTPYQIIKVDGSEKNGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTTVETKEASTPL
DVVILLDNSNSMSNIRHNHAHRAEKAGEATRALVDKITSNPDNRVALVTYGSTIFDGSEATVEKGVADANGKILN
DSALWTFDRTTFTAKTYNYSFLNLTSDPTDIQTIKDRIPSDAEELNKDKLMYQFGATFTQKALMTADDILTKQAR
PNSKKVIFHITDGVPTMSYPINFKYTGTTQSYRTQLNNFKAKTPNSSGILLEDFVTWSADGEHKIVRGDGESYQM
FTKKPVTDQYGVHQILSITSMEQRAKLVSAGYRFYGTDLYLYWRDSILAYPFNSSTDWITNHGDPTTWYYNGNMA
QDGYDVFTVGVGVNGDPGTDEATATRFMQSISSSPDNYTNVADPSQILQELNRYFYTIVNEKKSIENGTITDPMG
ELIDFQLGADGRFDPADYTLTANDGSSLVNNVPTGGPQNDGGLLKNAKVFYDTTEKRIRVTGLYLGTGEKVTLTY
NVRLNDQFVSNKFYDTNGRTTLHPKEVEKNTVRDFPIPKIRDVRKYPEITIPKEKKLGEIEFIKINKNDKKPLRD
AVFSLQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEV
RDVTSIVPQDIPAGYEFTNDKHYITNEPIPPKREYPRTGGIGMLPFYLIGCMMMGGVLLYTRKHP

Orf4_670 is a cell wall surface anchor family protein. An example of an amino acid sequence of orf4_670 is set forth in SEQ ID NO: 174.

SEQ ID NO: 174

MKSINKFLTMLAALLLTASSLFSAATVFAADNVSTAPDAVTKTLTIHKLLLSEDDLKTWDTNGPKGYDGTQSSLK
DLTGVVAEEIPNVYFELQKYNLTDGKEKENLKDDSKWTTVHGGLTTKDGLKIETSTLKGVYRIREDRTKTTYVGP
NGQVLTGSKAVPALVTLPLVNNNGTVIDAHVFPKNSYNKPVVDKRIADTLNYNDQNGLSIGTKIPYVVNTTIPSN
ATFATSFWSDEMTEGLTYNEDVTITLNNVAMDQADYEVTKGNNGFNLKLTEAGLAKINGKDADQKIQITYSATLN
SLAVADIPESNDITYHYGNHQDHGNTPKPTKPNNGQITVTKTWDSQPAPEGVKATVQLVNAKTGEKVGAPVELSE
NNWTYTWSGLDNSIEYKVEEEYNGYSAEYTVESKGKLGVKNWKDNNPAPINPEEPRVKTYGKKFVKVDQKDTRLE
NAQFVVKKADSNKYIAFKSTAQQAADEKAAATAKQKLDAAVAAYTNAADKQAAQALVDQAQQEYNVAYKEAKFGY
VEVAGKDEAMVLTSNTDGQFQISGLAAGTYKLEEIKAPEGFAKIDDVEFVVGAGSWNQGEFNYLKDVQKNDATKV
VNKKITIPQTGGIGTIIFAVAGAAIMGIAVYAYVKNNKDEDQLA

Orf5_670 is a cell wall surface anchor family protein. An example of an amino acid sequence of orf5_670 is set forth in SEQ ID NO: 175.

SEQ ID NO: 175

MTMQKMQKMISRIFFVMALCFSLVWGAHAVQAQEDHTLVLQLENYQEVVSQLPSRDGHRLQVWKLDDSYSYDDRV
QIVRDLHSWDENKLSSFKKTSFEMTFLENQIEVSHIPNGLYYVRSIIQTDAVSYPAEFLFEMTDQTVEPLVIVAK

KTDTMTTKVKETRVDQDHNRLEGVGFKLVSVARDGSEKEVPLIGEYRYSSSGQVGRTLYTDKNGEIFVTNLPLGN
YRFKEVEPLAGYAVTTLDTDVQLVDHQLVTITVVNQKLPRGNVDFMKVDGRTNTSLQGAMFKVMKEESGHYTPVL
QNGKEVVVTSGKDGRFRVEGLEYGTYYLWELQAPTGYVQLTSPVSFTIGKDTRKELVTVVKNNKRPRIDVPDTGE
ETLYILMLVAILLFGSGYYLTKKPNN

Orf6_670 is a sortase. An example of an amino acid sequence of orf6_670 is set forth in SEQ ID NO: 176.

SEQ ID NO: 176

MLIKMVKTKKQKRNNLLLGVVFFIGMAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDS
LNNVVSGDPWSEEMKKKGRAEYARMLEIHERMGHVEIPVIDVDLPVYAGTAEEVLQQGAGHLEGTSLPIGGNSTH
AVITAHTGLPTAKMFTDLTKLKVGDKFYVHNIKEVMAYQVDQVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINT
HRLLVRGHRIPYVAEVEEEFIAANKLSHLYRYLFYVAVGLIVILLWIIRRLRKKKKQPEKALKALKAARKEVKVE
DGQQ

Orf7_670 is a sortase. An example of an amino acid sequence of orf7_670 is set forth in SEQ ID NO: 177.

SEQ ID NO: 177

VSRYYYRIESNEVIKEFDETVSQMDKAELEERWRLAQAFNATLKPSEILDPFTEQEKKKGVSEYANMLKVHERIG
YVEIPAIDQEIPMYVGTSEEILQKGAGLLEGASLPVGGENTHTVVTAHRGLPTAELFSQLDKMKKGDVFYLHVLD
QVLAYQVDQILTVEPNDFEPVLIQHGEDYATLLTCTPYMINSHRLLVRGKRIPYTAPIAERNRAVRERGQFWLWL
LLAALVMILVLSYGVYRHRRIVKGLEKQLEEHHVKG

Orf8_670 is a sortase. An example of an amino acid sequence of orf8_670 is set forth in SEQ ID NO: 178.

SEQ ID NO: 178

MSKAKLQKLLGYLLMLVALVIPVYCFGQMVLQSLGQVKGHEIFSESVTADSYQEQLQRSLDYNQRLDSQNRIVDP
FLAEGYEVNYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLAMGLAHVDGTPLPVEGKGIRSVIAGHRAEPSH
VFFRHLDQLKVGDALYYDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAV
YQKSDPQTAAVARVAFTKEGQSVSRVATSQWLYRGLVVLAFLGILFVLWKLARLLRGK

As discussed above, a *S. pneumoniae* AI sequence is present in the 19A Hungary 6 *S. pneumoniae* genome. Examples of *S. pneumoniae* AI sequences from 19A Hungary 6 are set forth below.

ORF2_19AH is a transcriptional regulator. An example of an amino acid sequence of ORF2_19AH is set forth in SEQ ID NO: 187.

SEQ ID NO: 187

MLNKYIEKRITDKITILNILLDIRSIELDELSTLTSLQSKSLLSILQELQETFEEELTFNLDTQQVQLIEHHSHQ
TNYYFHQLYNQSTILKILRFFLLQGNQSFNEFTQKEYISIATGYRVRQKCGLLLRSVGLDLVKNQVVGPEYRIRF
LIALLQFHFGIEIYDLNDGSMDWVTHMIVQSNSQLSHELLEITPDEYVHFSILVALTWKRREFPLEFPESKEFEK
LKNLFMYPILMEHCQTYLEPHANMTFTQEELDYIFLVYCSANSSFSKDKWNQEKKTHTIQLILQHTRGKHLLSKF
KNILGNDISNSLSFLTALTFLTRTFLFGLQNLVPYYNYYEHYGIESDKPLYHISKAIVQEWMTEQKIEGVIDQHR
LYLFSLYLTETIFSSLPAIPIFIILNNQADVNLIKSIILRNFTDKVASVTGYNILISPPPSEEHLTEPLIIITTK
EYLPYVKKQYPKGKHHFLTIALDLHVSQQRLIYQTIVDIRKEAFDKRVAMIAKKAHYLL

ORF3_19AH is a cell wall surface protein. An example of an amino acid sequence of ORF3_19AH is set forth in SEQ ID NO: 188.

SEQ ID NO: 188

MKKVRKIFQKAVAGLCCISQLTAFSSIVALAETPETSPAIGKVVIKETGEGGALLGDAVFELKNNTDGTTVSQRT
EAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQWTVEVEKNGRTTVQGEQVENREEALSDQYPQTGTYPDVQTP
YQIIKVDGSEKNGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTTVETKEASTPLDVVILLD
NSNSMSNIRHNHAHRAEKAGEATRALVDKITSNPDNRVALVTYGSTIFDGSEATVEKGVADANGKILNDSALWTF
DRTTFTAKTYNYSFLNLTSDPTDIQTIKDRIPSDAEELNKDKLMYQFGATFTQKALMTADDILTKQARPNSKKVI
FHITDGVPTMSYPINFKYTGTTQSYRTQLNNFKAKTPNSSGILLEDFVTWSADGEHKIVRGDGESYQMFTKKPVT

DQYGVHQILSTTSMEQRAKLVSAGYRFYGTDLYLYWRDSILAYPFNSSTDWITNHGDPTTWYYNGNMAQDGYDVF
TVGVGVNGDPGTDEATATRFMQSISSSPDNYTNVADPSQILQELNRYFYTIVNEKKSIENGTITDPMGELIDFQL
GADGRFDPADYTLTANDGSSLVNNVPTGGPQNDGGLLKNAKVFYDTTEKRIRVTGLYLGTGEKVTLTYNVRLNDQ
FVSNKFYDTNGRTTLHPKEVEKNTVRDFPIPKIRDVRKYPEITIPKEKKLGEIEFIKINKNDKKPLRDAVFSLQK
QHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVTSIV
PQDIPAGYEFTNDKHYITNEPIPPKREYPRTGGIGMLPFYLIGCMMMGGVLLYTRKNP

ORF4_19AH is a cell wall surface protein. An example of an amino acid sequence of ORF4_19AH is set forth in SEQ ID NO: 189.

SEQ ID NO: 189

MKSINKFLTMLAALLLTASSLFSAATVFAADNVSTAPDAVTKTLTIHKLLLSEDDLKTWDTNGPKGYDGTQSSLK
DLTGVVAEEIPNVYFELQKYNLTDGKEKENLKDDSKWTTVHGGLTTKDGLKIETSTLKGVYRIREDRTKTTYVGP
NGQVLTGSKAVPALVTLPLVNNNGTVIDAHVFPKNSYNKPVVDKRIADTLNYNDQNGLSIGTKIPYVVNTTIPSN
ATFATSFWSDEMTEGLTYNEDVTITLNNVAMDQADYEVTKGXNGFNLKLTEAGLAKINGKDADQKIQITYSATLN
SLAVADIPESNDITYHYGNHQDHGNTPKPTKPNNGQITVTKTWDSQPAPEGVKATVQLVNAKTGEKVGAPVELSE
NNWTYTWSGLDNSIEYKVEEEYNGYSAEYTVESKGKLGVKNWKDNNPAPINPEEPRVKTYGKKFVKVDQKDTRLE
NAQFVVKKADSNKYIAFKSTAQQAADEKAAATAKQKLDAAVAAYTNAADKQAAQALVDQAQQEYNVAYKEAKFGY
VEVAGKDEAMVLTSNTDGQFQISGLAAGTYKLEEIKAPEGFAKIDDVEFVVGAGSWNQGEFNYLKDVQKNDATKV
VNKKITIPQTGGIGTIIFAVAGAAIMGIAVYAYVKNNKDEDQLA

ORF5_19AH is a cell wall surface protein. An example of an amino acid sequence of ORF5_19AH is set forth in SEQ ID NO: 190.

SEQ ID NO: 190

MTMQKMQKMISRIFFVMALCFSLVWGAHAVQAQEDHTLVLQLENYQEVVSQLPSRDGHRLQVWKLDDSYSYDDRV
QIVRDLHSWDENKLSSFKKTSFEMTFLENQIEVSHIPNGLYYVRSIIQTDAVSYPAEFLFEMTDQTVEPLVIVAK
KTDTMTTKVKLIKVDQDHNRLEGVGFKLVSVARDGSEKEVPLIGEYRYSSSGQVGRTLYTDKNGEIFVTNLPLGN
YRFKEVEPLAGYAVTTLDTDVQLVDHQLVTITVVNQKLPRGNVDFMKVDGRTNTSLQGAMFKVMKEESGHYTPVL
QNGKEVVVTSGKDGRFRVEGLEYGTYYLWELQAPTGYVQLTSPVSFTIGKDTRKELVTVVKNNKRPRIDVPDTGE
ETLYILMLVAILLFGSGYYLTKKPNN

ORF6_19AH is a putative sortase. An example of an amino acid sequence of ORF6_19AH is set forth in SEQ ID NO: 191.

SEQ ID NO: 191

MLIKMVKTKKQKRNNLLLGVVFFIGMAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDS
LNNVVSGDPWSEEMKKKGRAEYARMLEIHERMGHVEIPVIDVDLPVYAGTAEEVLQQGAGHLEGTSLPIGGNSTH
AVITAHTGLPTAKMFTDLTKLKVGDKFYVHNIKEVMAYQVDQVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINT
HRLLVRGHRIPYVAEVEEEFIAANKLSHLYRYLFYVAVGLIVILLWIIRRLRKKKKQPEKALKALKAARKEVKVE
DGQQ

ORF7_19AH is a putative sortase. An example of an amino acid sequence of ORF7_19AH is set forth in SEQ ID NO: 192.

SEQ ID NO: 192

MDNSRRSRKKGTKKKKHPLILLLLIFLVGFAVAIYPLVSRYYYRIESNEVIKEFDETVSQMDKAELEERWRLAQAF
NATLKPSEILDPFTEQEKKKGVSEYANMLKVHERIGYVEIPAIDQEIPMYVGTSEEILQKGAGLLEGASLPVGGE
NTHTVVTAHRGLPTAELFSQLDKMKKGDVFYLHVLDQVLAYQVDQILTVEPNDFEPVLIQHGEDYATLLTCTPYM
INSHRLLVRGKRIPYTAPIAERNRAVRERGQFWLWLLLAALVMILVLSYGVYRHRRIVKGLEKQLEEHHVKG

ORF8_19AH is a putative sortase. An example of an amino acid sequence of ORF8_19AH is set forth in SEQ ID NO: 193.

SEQ ID NO: 193

MSKAKLQKLLGYLLMLVALVIPVYCFGQMVLQSLGQVKGHEIFSESVTADSYQEQLQRSLDYNQRLDSQNRIVDP
FLAEGYEVNYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLAMGLAHVDGTPLPVEGKGIRSVIAGHRAEPSH
VFFRHLDQLKVGDALYYDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAV
YQKSDPQTAAVARVAFTKEGQSVSRVATSQWLYRGLVVLAFMGILFVLWKLARLLRGK

As discussed above, a *S. pneumoniae* AI sequence is present in the 6B Finland 12 *S. pneumoniae* genome. Examples of *S. pneumoniae* AI sequences from 6B Finland 12 are set forth below.

ORF2_6BF is a transcriptional regulator. An example of an amino acid sequence of ORF2_6BF is set forth in SEQ ID NO: 194.

SEQ ID NO: 194

MLNKYIEKRITDKITILNILLDIRSIELDELSTLTSLQSKSLLSILQELQETFEEELTFNLDTQQVQLIEHHSHQ
TNYYFHQLYNQSTILKILRFFLLQGNQSFNEFTQKEYISIATGYRVRQKCGLLLRSVGLDLVKNQVVGPEYRIRF
LIALLQFHFGIEIYDLNDGSMDWVTHMIVQSNSQLSHELLEITPDEYVHFSILVALTWKRREFPLEFPESKEFEK
LKNLFMYPILMEHCQTYLEPHANMTFTQEELDYIFLVYCSANSSFSKDKWNQEKKTHTIQLILQHTRGKHLLSKF
KNILGNDISNSLSFLTALTFLTRTFLFGLQNLVPYYNYYEHYGIESDKPLYHISKAIVQEWMTEQKIEGVIDQHR
LYLFSLYLTETIFSSLPAIPIFIILNNQADVNLIKSIILRNFTDKVASVTGYNILISPPPSEEHLTEPLIIITTK
EYLPYVKKQYPKGKHHFLTIALDLHVSQQRLIYQTIVDIRKEAFDKRVAMIAKKAHYLL

ORF3_6BF is a cell wall surface protein. An example of an amino acid sequence of ORF3_6BF is set forth in SEQ ID NO: 195.

SEQ ID NO: 195

MKKVRKIFQKAVAGLCCISQLTAFSSIVALAETPETSPAIGKVVIKETGEGGALLGDAVFELKNNTDGTTVSQRT
EAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQWTVEVEKNGRTTVQGEQVENREEALSDQYPQTGTYPDVQTP
YQIIKVDGSEKNGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTTVETKEASTPLDVVILLD
NSNSMSNIRHNHAHRAEKAGEATRALVDKITSNPDNRVALVTYGSTIFDGSEATVEKGVADANGKILNDSALWTF
DRTTFTAKTYNYSFLNLTSDPTDIQTIKDRIPSDAEELNKDKLMYQFGATFTQKALMTADDILTKQARPNSKKVI
FHITDGVPTMSYPINFKYTGTTQSYRTQLNNFKAKTPNSSGILLEDFVTWSADGEHKIVRGDGESYQMFTKKPVT
DQYGVHQILSITSMEQRAKLVSAGYRFYGTDLYLYWRDSILAYPFNSSTDWITNHGDPTTWYYNGNMAQDGYDVF
TVGVGVNGDPGTDEATATRFMQSISSSPDNYTNVADPSQILQELNRYFYTIVNEKKSIENGTITDPMGELIDFQL
GADGRFDPADYTLTANDGSSLVNNVPTGGPQNDGGLLKNAKVFYDTTEKRIRVTGLYLGTGEKVTLTYNVRLNDQ
FVSNKFYDTNGRTTLHPKEVEKNTVRDFPIPKIRDVRKYPEITIPKEKKLGEIEFIKINKNDKKPLRDAVFSLQK
QHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVTSIV
PQDIPAGYEFTNDKHYITNEPIPPKREYPRTGGIGMLPFYLIGCMMMGGVLLYTRKHP

ORF4_6BF is a cell wall surface protein. An example of an amino acid sequence of ORF4_6BF is set forth in SEQ ID NO: 196.

SEQ ID NO: 196

MKSINKFLTMLAALLLTASSLFSAATVFAADNVSTAPDAVTKTLTIHKLLLSEDDLKTWDTNGPKGYDGTQSSLK
DLTGVVAEEIPNVYFELQKYNLTDGKEKENLKDDSKWTTVHGGLTTKDGLKIETSTLKGVYRIREDRTKTTYVGP
NGQVLTGSKAVPALVTLPLVNNNGTVIDAHVFPKNSYNKPVVDKRIADTLNYNDQNGLSIGTKIPYVVNTTIPSN
ATFATSFWSDEMTEGLTYNEDVTITLNNVAMDQADYEVTKGNNGFNLKLTEAGLAKINGKDADQKIQITYSATLN
SLAVADIPESNDITYHYGNHQDHGNTPKPTKPNNGQITVTKTWDSQPAPEGVKATVQLVNAKTGEKVGAPVELSE
NNWTYTWSGLDNSIEYKVEEEYNGYSAEYTVESKGKLGVKNWKDNNPAPINPEEPRVKTYGKKFVKVDQKDTRLE
NAQFVVKKADSNKYIAFKSTAQQAADEKAAATAKQKLDAAVAAYTNAADKQAAQALVDQAQQEYNVAYKEAKFGY
VEVAGKDEAMVLTSNTDGQFQISGLAAGTYKLEEIKAPEGFAKIDDVEFVVGAGSWNQGEFNYLKDVQKNDATKV
VNKKITIPQTGGIGTIIFAVAGAAIMGIAVYAYVKNNKDEDQLA

ORF5_6BF is a cell wall surface protein. An example of an amino acid sequence of ORF5_6BF is set forth in SEQ ID NO: 197.

SEQ ID NO: 197

MTMQKMQKMISRIFFVMALCFSLVWGAHAVQAQEDHTLVLQLENYQEVVSQLPSRDGHRLQVWKLDDSYSYDDRV
QIVRDLHSWDENKLSSFKKTSFEMTFLENQIEVSHIPNGLYYVRSIIQTDAVSYPAEFLFEMTDQTVEPLVIVAK
KTDTMTTKVKLIKVDQDHNRLEGVGFKLVSVARDGSEKEVPLIGEYRYSSSGQVGRTLYTDKNGEIFVTNLPLGN
YRFKEVEPLAGYAVTTLDTDVQLVDHQLVTITVVNQKLPRGNVDFMKVDGRTNTSLQGAMFKVMKEESGHYTPVL
QNGKEVVVTSGKDGRFRVEGLEYGTYYLWELQAPTGYVQLTSPVSFTIGKDTRKELVTVVKNNKRPRIDVPDTGE
ETLYILMLVAILLFGSGYYLTKKPNN

ORF6_6BF is a putative sortase. An example of an amino acid sequence of ORF6_6BF is set forth in SEQ ID NO: 198.

SEQ ID NO: 198

MLIKMVKTKKQKRNNLLLGVVFFIGMAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDS
LNNVVSGDPWSEEMKKKGRAEYARMLEIHERMGHVEIPVIDVDLPVYAGTAEEVLQQGAGHLEGTSLPIGGNSTH
AVITAHTGLPTAKMFTDLTKLKVGDKFYVHNIKEVMAYQVDQVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINT
HRLLVRGHRIPYVAEVEEEFIAANKLSHLYRYLFYVAVGLIVILLWIIRRLRKKKKQPEKALKALKAARKEVKVE
DGQQ

ORF7_6BF is a putative sortase. An example of an amino acid sequence of ORF7_6BF is set forth in SEQ ID NO: 199.

SEQ ID NO: 199

MDNSRRSRKKGTKKKKHPLILLLIFLVGFAVAIYPLVSRYYYRIESNEVIKEFDETVSQMDKAELEERWRLAQAF
NATLKPSEILDPFTEQEKKKGVSEYANMLKVHERIGYVEIPAIDQEIPMYVGTSEEILQKGAGLLEGASLPVGGE
NTHTVVTAHRGLPTAELFSQLDKMKKGDVFYLHVLDQVLAYQVDQILTVEPNDFEPVLIQHGEDYATLLTCTPYM
INSHRLLVRGKRIPYTAPIAERNRAVRERGQFWLWLLLAALVMILVLSYGVYRHRRIVKGLEKQLEEHHVKG

ORF8_6BF is a putative sortase. An example of an amino acid sequence of ORF8_6BF is set forth in SEQ ID NO: 200.

SEQ ID NO: 200

MSKAKLQKLLGYLLMLVALVIPVYCFGQMVLQSLGQVKGHEIFSESVTADSYQEQLQRSLDYNQRLDSQNRIVDP
FLAEGYEVNYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLAMGLAHVDGTPLPVEGKGIRSVIAGHRAEPSH
VFFRHLDQLKVGDALYYDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAV
YQKSDPQTAAVARVAFTKEGQSVSRVATSQWLYRGLVVLAFLGILFVLWKLARLLRGK

As discussed above, a *S. pneumoniae* AI sequence is present in the 6B Spain 2 *S. pneumoniae* genome. Examples of *S. pneumoniae* AI sequences from 6B Spain 2 are set forth below.

ORF2_6BSP is a transcriptional regulator. An example of an amino acid sequence of ORF2_6BSP is set forth in SEQ ID NO: 201.

SEQ ID NO: 201

MLNKYIEKRITDKITILNILLDIRSIELDELSTLTSLQSKSLLSILQELQETFEEELTFNLDTQQVQLIEHHSHQ
TNYYFHQLYNQSTILKILRFFLLQGNQSFNEFTQKEYISIATGYRVRQKCGLLLRSVGLDLVKNQVVGPEYRIRF
LIALLQFHFGIEIYDLNDGSMDWVTHMIVQSNSQLSHELLEITPDEYVHFSILVALTWKRREFPLEFPESKEFEK
LKNLFMYPILMEHCQTYLEPHANMTFTQEELDYIFLVYCSANSSFSKDKWNQEKKTHTIQLILQHTRGKHLLSKF
KNILGNDISNSLSFLTALTFLTRTFLFGLQNLVPYYNYYEHYGIESDKPLYHISKAIVQEWMTEQKIEGVIDQHR
LYLFSLYLTETIFSSLPAIPIFIILNNQADVNLIKSIILRNFTDKVASVTGYNILISPPPSEEHLTEPLIIITTK
EYLPYVKKQYPKGKHHFLTIALDLHVSQQRLIYQTIVDIRKEAFDKRVAMIAKKAHYLL

ORF3_6BSP is a cell wall surface protein. An example of an amino acid sequence of ORF3_6BSP is set forth in SEQ ID NO: 202.

SEQ ID NO: 202

MKKVRKIFQKAVAGLCCISQLTAFSSIVALAETPETSPAIGKVVIKETGEGGALLGDAVFELKNNTDGTTVSQRT
EAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQWTVEVEKNGRTTVQGEQVENREEALSDQYPQTGTYPDVQTP
YQIIKVDGSEKNGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTTVETKEASTPLDVVILLD
NSNSMSNIRHNHAHRAEKAGEATRALVDKITSNPDNRVALVTYGSTIFDGSEATVEKGVADANGKILNDSALWTF
DRTTFTAKTYNYSFLNLTSDPTDIQTIKDRIPSDAEELNKDKLMYQFGATFTQKALMTADDILTKQARPNSKKVI
FHITDGVPTMSYPINFKYTGTTQSYRTQLNNFKAKTPNSSGILLEDFVTWSADGEHKIVRGDGESYQMFTKKPVT
DQYGVHQILSITSMEQRAKLVSAGYRFYGTDLYLYWRDSILAYPFNSSTDWITNHGDPTTWYYNGNMAQDGYDVF
TVGVGVNGDPGTDEATATRFMQSISSSPDNYTNVADPSQILQELNRYFYTIVNEKKSIENGTITDPMGELIDFQL
GADGRFDPADYTLTANDGSSLVNNVPTGGPQNDGGLLKNAKVFYDTTEKRIRVTGLYLGTGEKVTLTYNVRLNDQ
FVSNKFYDTNGRTTLHPKEVEKNTVRDFPIPKIRDVRKYPEITIPKEKKLGEIEFIKINKNDKKPLRDAVFSLQK
QHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVTSIV
PQDIPAGYEFTNDKHYITNEPIPPKREYPRTGGIGMLPFYLIGCMMMGGVLLYTRKHP

ORF4_6BSP is a cell wall surface protein. An example of an amino acid sequence of ORF4_6BSP is set forth in SEQ ID NO: 203.

SEQ ID NO: 203

MKSINKFLTMLAALLLTASSLFSAATVFAADNVSTAPDAVTKTLTIHKLLLSEDDLKTWDTNGPKGYDGTQSSLK
DLTGVVAEEIPNVYFELQKYNLTDGKEKENLKDDSKWTTVHGGLTTKDGLKIETSTLKGVYRIREDRTKTTYVGP
NGQVLTGSKAVPALVTLPLVNNNGTVIDAHVFPKNSYNKPVVDKRIADTLNYNDQNGLSIGTKIPYVVNTTIPSN
ATFATSFWSDEMTEGLTYNEDVTITLNNVAMDQADYEVTKGNNGFNLKLTEAGLAKINGKDADQKIQITYSATLN
SLAVADIPESNDITYHYGNHQDHGNTPKPTKPNNGQITVTKTWDSQPAPEGVKATVQLVNAKTGEKVGAPVELSE
NNWTYTWSGLDNSIEYKVEEEYNGYSAEYTVESKGKLGVKNWKDNNPAPINPEEPRVKTYGKKFVKVDQKDTRLE
NAQFVVKKADSNKYIAFKSTAQQAADEKAAATAKQKLDAAVAAYTNAADKQAAQALVDQAQQEYNVAYKEAKFGY
VEVAGKDEAMVLTSNTDGQFQISGLAAGTYKLEEIKAPEGFAKIDDVEFVVGAGSWNQGEFNYLKDVQKNDATKV
VNKKITIPQTGGIGTIIFAVAGAAIMGIAVYAYVKNNKDEDQLA

ORF5_6BSP is a cell wall surface protein. An example of an amino acid sequence of ORF5_6BSP is set forth in SEQ ID NO: 204.

SEQ ID NO: 204

MTMQKMQKMISRIFFVMALCFSLVWGAHAVQAQEDHTLVLQLENYQEVVSQLPSRDGHRLQVWKLDDSYSYDDRV
QIVRDLHSWDENKLSSFKKTSFEMTFLENQIEVSHIPNGLYYVRSIIQTDAVSYPAEFLFEMTDQTVEPLVIVAK
KTDTMTTKVKLIKVDQDHNRLEGVGFKLVSVARDGSEKEVPLIGEYRYSSSGQVGRTLYTDKNGEIFVTNLPLGN
YRFKEVEPLAGYAVTTLDTDVQLVDHQLVTITVVNQKLPRGNVDFMKVDGRTNTSLQGAMFKVMKEESGHYTPVL
QNGKEVVVTSGKDGRFRVEGLEYGTYYLWELQAPTGYVQLTSPVSFTIGKDTRKELVTVVKNNKRPRIDVPDTGE
ETLYILMLVAILLFGSGYYLTKKPNN

ORF6_6BSP is a putative sortase. An example of an amino acid sequence of ORF6_6BSP is set forth in SEQ ID NO: 205.

SEQ ID NO: 205

MLIKMVKTKKQKRNNLLLGVVFFIGMAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDS
LNNVVSGDPWSEEMKKKGRAEYARMLEIHERMGHVEIPVIDVDLPVYAGTAEEVLQQGAGHLEGTSLPIGGNSTH
AVITAHTGLPTAKMFTDLTKLKVGDKFYVHNIKEVMAYQVDQVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINT
HRLLVRGHRIPYVAEVEEEFIAANKLSHLYRYLFYVAVGLIVILLWIIRRLRKKKKQPEKALKALKAARKEVKVE
DGQQ

ORF7_6BSP is a putative sortase. An example of an amino acid sequence of ORF7_6BSP is set forth in SEQ ID NO: 206.

SEQ ID NO: 206

MDNSRRSRKKGTKKKKHPLILLLLIFLVGFAVAIYPLVSRYYYRIESNEVIKEFDETVSQMDKAELEERWRLAQAF
NATLKPSEILDPFTEQEKKKGVSEYANMLKVHERIGYVEIPAIDQEIPMYVGTSEEILQKGAGLLEGASLPVGGE
NTHTVVTAHRGLPTAELFSQLDKMKKGDVFYLHVLDQVLAYQVDQILTVEPNDFEPVLIQHGEDYATLLTCTPYM
INSHRLLVRGKRIPYTAPIAERNRAVRERGQFWLWLLLAALVMILVLSYGVYRHRRIVKGLEKQLEEHHVKG

ORF8_6BSP is a putative sortase. An example of an amino acid sequence of ORF8_6BSP is set forth in SEQ ID NO: 207.

SEQ ID NO: 207

MSKAKLQKLLGYLLMLVALVIPVYCFGQMVLQSLGQVKGHEIFSESVTADSYQEQLQRSLDYNQRLDSQNRIVDP
FLAEGYEVNYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLAMGLAHVDGTPLPVEGKGIRSVIAGHRAEPSH
VFFRHLDQLKVGDALYYDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAV
YQKSDPQTAAVARVAFTKEGQSVSRVATSQWLYRGLVVLAFLGILFVLWKLARLLRGK

As discussed above, a *S. pneumoniae* AI sequence is present in the 9V Spain 3 *S. pneumoniae* genome. Examples of *S. pneumoniae* AI sequences from 9V Spain 3 are set forth below.

ORF2_9VSP is a transcriptional regulator. An example of an amino acid sequence of ORF2_9VSP is set forth in SEQ ID NO: 208.

PCT/US05/27239

SEQ ID NO: 208

MLNKYIEKRITDKITILNILLDIRSIELDELSTLTSLQSKSLLSILQELQETFEEELTFNLDTQQVQLIEHHSHQ
TNYYFHQLYNQSTILKILRFFLLQGNQSFNEFTQKEYISIATGYRVRQKCGLLLRSVGLDLVKNQVVGPEYRIRF
LIALLQFHFGIEIYDLNDGSMDWVTHMIVQSNSQLSHELLEITPDEYVHFSILVALTWKRREFPLEFPESKEFEK
LKNLFMYPILMEHCQTYLEPHANMTFTQEELDYIFLVYCSANSSFSKDKWNQEKKTHTIQLILQHTRGKHLLSKF
KNILGNDISNSLSFLTALTFLTRTFLFGLQNLVPYYNYYEHYGIESDKPLYHISKAIVQEWMTEQKIEGVIDQHR
LYLFSLYLTETIFSSLPAIPIFIILNNQADVNLIKSIILRNFTDKVASVTGYNILISPPPSEEHLTEPLIIITTK
EYLPYVKKQYPKGKHHFLTIALDLHVSQQRLIYQTIVDIRKEAFDKRVAMIAKKAHYLL

ORF3_9VSP is a cell wall surface protein. An example of an amino acid sequence of ORF3_9VSP is set forth in SEQ ID NO: 209.

SEQ ID NO: 209

MKKVRKIFQKAVAGLCCISQLTAFSSIVALAETPETSPAIGKVVIKETGEGGALLGDAVFELKNNTNGTTVSQRT
EAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQRTVEVEKNGRTTVQGEQVENREEALSDQYPQTGTYPDVQTP
YQIIKVDGSEKNGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTVYERKDKSVPLDVVILLD
NSNSMSNIRNKNARRAERAGEATRSLIDKITSDPENRVALVTYASTIFDGTEFTVEKGVADKNGKRLNDSLFWNY
DQTSFTTNTKDYSYLKLTNDKNDIVELKNKVPTEAEDHDGNRLMYQFGATFTQKALMKADEILTQQARQNSQKVI
FHITDGVPTMSYPINFNHATFAPSYQNQLNAFFSKSPNKDGILLSDFITQATSGEHTIVRGDGQSYQMFTDKTVY
EKGAPAAFPVKPEKYSEMKAVGYAVIGDPINGGYIWLNWRESILAYPFNSNTAKITNHGDPTRWYYNGNIAPDGY
DVFTVGIGINGDPGTDEATATSFMQSISSKPENYTNVTDTTKILEQLNRYFHTIVTEKKSIENGTITDPMGELID
LQLGTDGRFDPADYTLTANDGSRLENGQAVGGPQNDGGLLKNAKVFYDTTEKRIRVTGLYLGTGEKVTLTYNVRL
NDQFVSNKFYDTNGRTTLHPKEVEKNTVRDFPIPKIRDVRKYPAITIAKEKKLGEIEFIKINKNDKKPLRDAVFS
LQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVT
SIVPQDIPAGYEFTNDKHYITNEPIPPKREYPRTGGIGMLLFYLIGCMMMGGVLLYTRKHP

ORF4_9VSP is a cell wall surface protein. An example of an amino acid sequence of ORF4_9VSP is set forth in SEQ ID NO: 210.

SEQ ID NO: 210

MKSINKFLTMLAALLLTASSLFSAATVFAAGTTTTSVTVHKLLATDGDMDKIANELETGNYAGNKVGVLPANAKE
IAGVMFVWTNTNNEIIDENGQTLGVNIDPQTFKLSGAMPATAMKKLTEAEGAKFNTANLPAAKYKIYEIHSLSTY
VGEDGATLTGSKAVPIEIELPLNDVVDAHVYPKNTEAKPKIDKDFKGKANPDTPRVDKDTPVNHQVGDVVEYEIV
TKIPALANYATANWSDRMTEGLAFNKGTVKVTVDDVALEAGDYALTEVATGFDLKLTDAGLAKVNDQNAEKTVKI
TYSATLNDKAIVEVPESNDVTFNYGNNPDHGNTPKPNKPNENGDLTLTKTWVDATGAPIPAGAEATFDLVNAQTG
KVVQTVTLTTDKNTVTVNGLDKNTEYKFVERSIKGYSADYQEITTAGEIAVKNWKDENPKPLDPTEPKVVTYGKK
FVKVNDKDNRLAGAEFVIANADNAGQYLARKADKVSQEEKQLVVTTKDALDRAVAAYNALTAQQQTQQEKEKVDK
AQAAYNAAVIAANNAFEWVADKDNENVVKLVSDAQGRFEITGLLAGTYYLEETKQPAGYALLTSRQKFEVTATSY
SATGQGIEYTAGSGKDDATKVVNKKITIPQTGGIGTIIFAVAGAVIMGIAVYAYVKNNKDEDQLA

ORF5_9VSP is a cell wall surface protein. An example of an amino acid sequence of ORF5_9VSP is set forth in SEQ ID NO: 211.

SEQ ID NO: 211

MTMQKMQKMQKMQKMQKMQKMQKMISRIFFVMALCFSLVWGAHAVQAQEDHTLVLQLENYQEVVSQLPSRDGHRLQVW
KLDDSYSYDNRVQIVRDLHSWDENKLSSFKKTSFEMTFLENQIEVSHIPNGLYYVRSIIQTDAVSYPAEFLFEMT
DQTVEPLVIVAKKADTVTTKVKLIKVDQDHNRLEGVGFKLVSVARDGSEKEVPLIGEYRYSSSGQVGRTLYTDKN
GEIVVTNLPLGTYRFKEVEPLAGYTVTTMDTDVQLVDHQLVTITVVNQKLPRGNVDFMKVDGRTNTSLQGAMFKV
MKEENGHYTPVLQNGKEVVVASGKDGRFRVEGLEYGTYYLWELQAPTGYVQLTSPVSFTIGKDTRKELVTVVKNN
KRPRIDVPDTGEETLYILMLVAILLFGSGYYLTKKTNN

ORF6_9VSP is a putative sortase. An example of an amino acid sequence of ORF6_9VSP is set forth in SEQ ID NO: 212.

SEQ ID NO: 212

MLIKMAKTKKQKRNNLLLGVVFFIGIAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDS
LNNVVSGDPWSEEMKKKGRAEYARMLEIHERMGHVEIPAIDVDLPVYAGTAEEVLQQGAGHLEGTSLPIGGNSTH
AVITAHTGLPTAKMFTDLTKLKVGDKFYVHNIKEVMAYQVDQVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINT
HRLLVRGHRIPYVAEVEEEFIAANKLSHLYRYLFYVAVGLIVILLWIIRRLRKKKRQSERALKALKEATKEVKVE
DE

PCT/US05/27239

ORF7_9VSP is a putative sortase. An example of an amino acid sequence of ORF7_9VSP is set forth in SEQ ID NO: 213.

SEQ ID NO: 213

MSKSRYSRKKSVKKKKNPFILLLIFLVGLAVAMYPLVSRYYYRIESNEVIKEFDETVSQMDKAELEERWRLAQAF
NATLKPSEILDPFTEQEKKKGVSEYANMLKVHERIGYVEIPAIDQEIPMYVGTSEEILQKGAGLLEGASLPVGGE
NTHTVVTAHRGLPTAELFSQLDKMKKGDIFYLHVLDQVLAYQVDQIVTVEPNDFEPVLIQHGEDYATLLTCTPYM
INSHRLLVRGKRIPYTAPIAERNRAVRERGQFWLWLLLGAMAVILLLLYRVYRNRRIVKGLEKQLEGRHVKD

ORF8_9VSP is a putative sortase. An example of an amino acid sequence of ORF8_9VSP is set forth in SEQ ID NO: 214.

SEQ ID NO: 214

MSRTKLRALLGYLLMLVACLIPIYCFGQMVLQSLGQVKGHATFVKSMTTEMYQEQQNHSLAYNQRLASQNRIVDP
FLAEGYEVNYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLGMGLAHVDGTPLPLDGTGIRSVIAGHRAEPSH
VFFRHLDQLKVGDALYYDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAV
YQKSDPQTAAVARVAFTKEGQSVSRVATSQWLYRGLVVLAFLGILFVLWKLARLLRGK

As discussed above, a *S. pneumoniae* AI sequence is present in the 14 CSR 10 *S. pneumoniae* genome. Examples of *S. pneumoniae* AI sequences from 14 CSR 10 are set forth below.

ORF2_14CSR is a transcriptional regulator. An example of an amino acid sequence of ORF2_14CSR is set forth in SEQ ID NO: 215.

SEQ ID NO: 215

MLNKYIEKRITDKITILNILLDIRSIELDELSTLTSLQSKSLLSILQELQETFEEELTFNLDTQQVQLIEHHSHQ
TNYYFHQLYNQSTILKILRFFLLQGNQSFNEFTQKEYISIATGYRVRQKCGLLLRSVGLDLVKNQVVGPEYRIRF
LIALLQFHFGIEIYDLNDGSMDWVTHMIVQSNSQLSHELLEITPDEYVHFSILVALTWKRREFPLEFPESKEFEK
LKNLFMYPILMEHCQTYLEPHANMTFTQEELDYIFLVYCSANSSFSKDKWNQEKKTHTIQLILQHTRGKHLLSKF
KNILGNDISNSLSFLTALTFLTRTFLFGLQNLVPYYNYYEHYGIESDKPLYHISKAIVQEWMTEQKIEGVIDQHR
LYLFSLYLTETIFSSLPAIPIFIILNNQADVNLIKSIILRNFTDKVASVTGYNILISPPPSEEHLTEPLIIITTK
EYLPYVKKQYPKGKHHFLTIALDLHVSQQRLIYQTIVDIRKEAFDKRVAMIAKKAHYLL

ORF3_14CSR is a cell wall surface protein. An example of an amino acid sequence of ORF3_14CSR is set forth in SEQ ID NO: 216.

SEQ ID NO: 216

MKKVRKIFQKAVAGLCCISQLTAFSSIVALAETPETSPAIGKVVIKETGEGGALLGDAVFELKNNTDGTTVSQRT
EAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQWTVEVEKNGRTTVQGEQVENREEALSDQYPQTGTYPDVQTP
YQIIKVDGSEKNGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTTVETKEASTPLDVVILLD
NSNSMSNIRHNHAHRAEKAGEATRALVDKITSNPDNRVALVTYGSTIFDGSEATVEKGVADANGKILNDSALWTF
DRTTFTAKTYNYSFLNLTSDPTDIQTIKDRIPSDAEELNKDKLMYQFGATFTQKALMTADDILTKQARPNSKKVI
FHITDGVPTMSYPINFKYTGTTQSYRTQLNNFKAKTPNSSGILLEDFVTWSADGEHKIVRGDGESYQMFTKKPVT
DQYGVHQILSITSMEQRAKLVSAGYRFYGTDLYLYWRDSILAYPFNSSTDWITNHGDPTTWYYNGNMAQDGYDVF
TVGVGVNGDPGTDEATATRFMQSISSSPDNYTNVADPSQILQELNRYFYTIVNEKKSIENGTITDPMGELIDFQL
GADGRFDPADYTLTANDGSSLVNNVPTGGPQNDGGLLKNAKVFYDTTEKRIRVTGLYLGTGEKVTLTYNVRLNDQ
FVSNKFYDTNGRTTLHPKEVEKNTVRDFPIPKIRDVRKYPEITIPKEKKLGEIEFIKINKNDKKPLRDAVFSLQK
QHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVTSIV
PQDIPAGYEFTNDKHYITNEPIPPKREYPRTGGIGMLPFYLIGCMMMGGVLLYTRKHP

ORF4_14CSR is a cell wall surface protein. An example of an amino acid sequence of ORF4_14CSR is set forth in SEQ ID NO: 217.

SEQ ID NO: 217

MKSINKFLTMLAALLLTASSLFSAATVFAADNVSTAPDAVTKTLTIHKLLLSEDDLKTWDTNGPKGYDGTQSSLK
DLTGVVAEEIPNVYFELQKYNLTDGKEKENLKDDSKWTTVHGGLTTKDGLKIETSTLKGVYRIREDRTKTTYVGP
NGQVLTGSKAVPALVTLPLVNNNGTVIDAHVFPKNSYNKPVVDKRIADTLNYNDQNGLSIGTKIPYVVNTTIPSN
ATFATSFWSDEMTEGLTYNEDVTITLNNVAMDQADYEVTKGNNGFNLKLTEAGLAKINGKDADQKIQITYSATLN
SLAVADIPESNDITYHYGNHQDHGNTPKPTKPNNGQITVTKTWDSQPAPEGVKATVQLVNAKTGEKVGAPVELSE

NNWEYTWSGLDNSTEIKVEEEYNGYEAEYTVESKGKLGVKNWKDNNPAPINPEEPRVKTYGKKFVKVDQKDTRLE
NAQFVVKKADSNKYIAFKSTAQQAADEKAAATAKQKLDAAVAAYTNAADKQAAQALVDQAQQEYNVAYKEAKFGY
VEVAGKDEAMVLTSNTDGQFQISGLAAGTYKLEEIKAPEGFAKIDDVEFVVGAGSWNQGEFNYLKDVQKNDATKV
VNKKITIPQTGGIGTIIFAVAGAAIMGIAVYAYVKNNKDEDQLA

ORF5_14CSR is a cell wall surface protein. An example of an amino acid sequence of ORF5_14CSR is set forth in SEQ ID NO: 218.

SEQ ID NO: 218

MTMQKMQKMISRIFFVMALCFSLVWGAHAVQAQEDHTLVLQLENYQEVVSQLPSRDGHRLQVWKLDDSYSYDDRV
QIVRDLHSWDENKLSSFKKTSFEMTFLENQIEVSHIPNGLYYVRSIIQTDAVSYPAEFLFEMTDQTVEPLVIVAK
KTDTMTTKVKLIKVDQDHNRLEGVGFKLVSVARDGSEKEVPLIGEYRYSSSGQVGRTLYTDKNGEIFVTNLPLGN
YRFKEVEPLAGYAVTTLDTDVQLVDHQLVTITVVNQKLPRGNVDFMKVDGRTNTSLQGAMFKVMKEESGHYTPVL
QNGKEVVVTSGKDGRFRVEGLEYGTYYLWELQAPTGYVQLTSPVSFTIGKDTRKELVTVVKNNKRPRIDVPDTGE
ETLYILMLVAILLFGSGYYLTKKPNN

ORF6_14CSR is a putative sortase. An example of an amino acid sequence of ORF6_14CSR is set forth in SEQ ID NO: 219.

SEQ ID NO: 219

MLIKMVKTKKQKRNNLLLGVVFFIGMAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDS
LNNVVSGDPWSEEMKKKGRAEYARMLEIHERMGHVEIPVIDVDLPVYAGTAEEVLQQGAGHLEGTSLPIGGNSTH
AVITAHTGLPTAKMFTDLTKLKVGDKFYVHNIKEVMAYQVDQVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINT
HRLLVRGHRIPYVAEVEEEFIAANKLSHLYRYLFYVAVGLIVILLWIIRRLRKKKKQPEKALKALKAARKEVKVE
DGQQ

ORF7_14CSR is a putative sortase. An example of an amino acid sequence of ORF7_14CSR is set forth in SEQ ID NO: 220.

SEQ ID NO: 220

MDNSRRSRKKGTKKKKHPLILLLIFLVGFAVAIYPLVSRYYYRIESNEVIKEFDETVSQMDKAELEERWRLAQAF
NATLKPSEILDPFTEQEKKKGVSEYANMLKVHERIGYVEIPAIDQEIPMYVGTSEEILQKGAGLLEGASLPVGGE
NTHTVVTAHRGLPTAELFSQLDKMKKGDVFYLHVLDQVLAYQVDQILTVEPNDFEPVLIQHGEDYATLLTCTPYM
INSHRLLVRGKRIPYTAPIAERNRAVRERGQFWLWLLLAALVMILVLSYGVYRHRRIVKGLEKQLEEHHVKG

ORF8_14CSR is a putative sortase. An example of an amino acid sequence of ORF8_14CSR is set forth in SEQ ID NO: 221.

SEQ ID NO: 221

MSKAKLQKLLGYLLMLVALVIPVYCFGQMVLQSLGQVKGHEIFSESVTADSYQEQLQRSLDYNQRLDSQNRIVDP
FLAEGYEVNYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLAMGLAHVDGTPLPVEGKGIRSVIAGHRAEPSH
VFFRHLDQLKVGDALYYDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAV
YQKSDPQTAAVARVAFTKEGQSVSRVATSQWLYRGLVVLAFLGILFVLWKLARLLRGK

As discussed above, a *S. pneumoniae* AI sequence is present in the 19F Taiwan 14 *S. pneumoniae* genome. Examples of *S. pneumoniae* AI sequences from 19F Taiwan 14 are set forth below.

ORF2_19FTW is a transcriptional regulator. An example of an amino acid sequence of ORF2_19FTW is set forth in SEQ ID NO: 222.

SEQ ID NO: 222

MLNKYIEKRITDKITILNILLDIRSIELDELSTLTSLQSKSLLSILQELQETFEEELTFNLDTQQVQLIEHHSHQ
TNYYFHQLYNQSTILKILRFFLLQGNQSFNEFTQKEYISIATGYRVRQKCGLLLRSVGLDLVKNQVVGPEYRIRF
LIALLQFHFGIEIYDLNDGSMDWVTHMIVQSNSQLSHELLEITPDEYVHFSILVALTWKRREFPLEFPESKEFEK
LKNLFMYPILMEHCQTYLEPHANMTFTQEELDYIFLVYCSANSSFSKDKWNQEKKTHTIQLILQHTRGKHLLSKF
KNILGNDISNSLSFLTALTFLTRTFLFGLQNLVPYYNYYEHYGIESDKPLYHISKAIVQEWMTEQKIEGVIDQHR
LYLFSLYLTETIFSSLPAIPIFIILNNQADVNLIKSIILRNFTDKVASVTGYNILISPPPSEEHLTEPLIIITTK
EYLPYVKKQYPKGKHHFLTIALDLHVSQQRLIYQTIVDIRKEAFDKRVAMIAKKAHYLL

PCT/US05/27239

ORF3_19FTW is a cell wall surface protein. An example of an amino acid sequence of ORF3_19FTW is set forth in SEQ ID NO: 223.

SEQ ID NO: 223

MKKVRKIFQKAVAGLCCISQLTAFSSIVALAETPETSPAIGKVVIKETGEGGALLGDAVFELKNNTDGTTVSQRT
EAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQWTVEVEKNGRTTVQGEQVENREEALSDQYPQTGTYPDVQTP
YQIIKVDGSEKNGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTVYERKDKSVPLDVVILLD
NSNSMSNIRNKNARRAERAGEATRSLIDKITSDPENRVALVTYASTIFDGTEFTVEKGVADKNGKRLNDSLFWNY
DQTSFTTNTKDYSYLKLTNDKNDIVELKNKVPTEAEDHDGNRLMYQFGATFTQKALMKADEILTQQARQNSQKVI
FHITDGVPTMSYPINFNHATFAPSYQNQLNAFFSKSPNKDGILLSDFITQATSGEHTIVRGDGQSYQMFTDKTVY
EKGAPAAFPVKPEKYSEMKAVGYAVIGDPINGGYIWLNWRESILAYPFNSNTAKITNHGAPTRWYYNGNIAPDGY
DVFTVGIGINGDPGTDEATATSFMQSISSKPENYTNVTDTTKILEQLNRYFHTIVTEKKSIENGTITDPMGELID
LQLGTDGRFDPADYTLTANDGSRLENGQAVGGPQNDGGLLKNAKVFYDTTEKRIRVTGLYLGTGEKVTLTYNVRL
NDQFVSNKFYDTNGRTTLHPKEVEKNTVRDFPIPKIRDVRKYPAITIAKEKKLGEIEFIKINKNDKKPLRDAVFS
LQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVT
SIVPQDIPAGYEFTNDKHYITNEPIPPKREYPRTGGIGMLPFYLIGCMMMGGVLLYTRKHP

ORF4_19FTW is a cell wall surface protein. An example of an amino acid sequence of ORF4_19FTW is set forth in SEQ ID NO: 224.

SEQ ID NO: 224

MKSINKFLTMLAALLLTASSLFSAATVFAAGTTTTSVTVHKLLATDGDMDKIANELETGNYAGNKVGVLPANAKE
IAGVMFVWTNTNNEIIDENGQTLGVNIDPQTFKLSGAMPATAMKKLTEAEGAKFNTANLPAAKYKIYEIHSLSTY
VGEDGATLTGSKAVPIEIELPLNDVVDAHVYPKNTEAKPKIDKDFKGKANPDTPRVDKDTPVNHQVGDVVEYEIV
TKIPALANYATANWSDRMTEGLAFNKGTVKVTVDDVALEAGDYALTEVATGFDLKLTDAGLAKVNDQNAEKTVKI
TYSATLNDKAIVEVPESNDVTFNYGNNPDHGNTPKPNKPNENGDLTLTKTWVDATGAPIPAGAEATFDLVNAQTG
KVVQTVTLTTDKNTVTVNGLDKNTEYKFVERSIKGYSADYQEITTAGEIAVKNWKDENPKPLDPTEPKVVTYGKK
FVKVNDKDNRLAGAEFVIANADNAGQYLARKADKVSQEEKQLVVTTKDALDRAVAAYNALTAQQQTQQEKEKVDK
AQAAYNAAVIAANNAFEWVADKDNENVVKLVSDAQGRFEITGLLAGTYYLEETKQPAGYALLTSRQKFEVTATSY
SATGQGIEYTAGSGKDDATKVVNKKITIPQTGGIGTIIFAVAGAVIMGIAVYAYVKNNKDEDQLA

ORF5_19FTW is a cell wall surface protein. An example of an amino acid sequence of ORF5_19FTW is set forth in SEQ ID NO: 225.

SEQ ID NO: 225

MTMQKMQKMISRIFFVMALCFSLVWGAHAVQAQEDHTLVLQLENYQEVVSQLPSRDGHRLQVWKLDDSYSYDNRV
QIVRDLHSWDENKLSSFKKTSFEMTFLENQIEVSHIPNGLYYVRSIIQTDAVSYPAEFLFEMTDQTVEPLVIVAK
KADTVTTKVKLIKVDQDHNRLEGVGFKLVSVARDGSEKEVPLIGEYRYSSSGQVGRTLYTDKNGEIVVTNLPLGT
YRFKEVEPLAGYTVTTMDTDVQLVDHQLVTITVVNQKLPRGNVDFMKVDGRTNTSLQGAMFKVMKEENGHYTPVL
QNGKEVVVASGKDGRFRVEGLEYGTYYLWELQAPTGYVQLTSPVSFTIGKDTRKELVTVVKNNKRPRIDVPDTGE
ETLYILMLVAILLFGSGYYLTKKTNN

ORF6_19FTW is a putative sortase. An example of an amino acid sequence of ORF6_19FTW is set forth in SEQ ID NO: 226.

SEQ ID NO: 226

MLIKMAKTKKQKRNNLLLGVVFFIGMAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDS
LNNVVSGDPWSEEMKKKGRAEYARMLEIHERMGHVEIPAIDVDLPVYAGTAEEVLQQGAGHLEGTSLPIGGNSTH
AVITAHTGLPTAKMFTDLTKLKVGDKFYVHNIKEVMAYQVDQVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINT
HRLLVRGHRIPYVAEVEEEFIAANKLSHLYRYLFYVAVGLIVILLWIIRRLRKKKRQSERALKALKEATKEVKVE
DE

ORF7_19FTW is a putative sortase. An example of an amino acid sequence of ORF7_19FTW is set forth in SEQ ID NO: 227.

SEQ ID NO: 227

MSKSRYSRKKSVKKKKNPFILLLIFLVGLAVAMYPLVSRYYYRIESNEVIKEFDETVSQMDKAELEERWRLAQAF
NATLKPSEILDPFTDQEKKQGVSEYANMLKVHERIGYVEIPAIEQEIPMYVGTSEDILQKGAGLLEGASLPVGGE

NPHCVITAHRGLETAELFSQIDKMKKGDIFFLHVLDQVLAYQVDQIVTVEPNDFEPVLIQHGQDYATLLTCTPYM
INSHRLLVRGKRIPYTAPIAERNRAVRERGQFWLWLLLGAMAVILLLLYRVYRNRRIVKGLEKQLEGRHVKD

ORF8_19FTW is a putative sortase. An example of an amino acid sequence of ORF8_19FTW is set forth in SEQ ID NO: 228.

SEQ ID NO: 228

MSRTKLRALLGYLLMLVACLIPIYCFGQMVLQSLGQVKGHATFVKSMTTEMYQEQQNHSLAYNQRLASQNRIVDP
FLAEGYEVNYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLGMGLAHVDGTPLPLDGTGIRSVIAGHRAEPSH
VFFRHLDQLKVGDALYYDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAV
YQKSDPQTAAVARVAFTKEGQSVSRVATSQWLYRGLVVLAFLGILFVLWKLARLLRGK

As discussed above, a *S. pneumoniae* AI sequence is present in the 23F Taiwan 15 *S. pneumoniae* genome. Examples of *S. pneumoniae* AI sequences from 23F Taiwan 15 are set forth below.

ORF2_23FTW is a transcriptional regulator. An example of an amino acid sequence of ORF2_23FTW is set forth in SEQ ID NO: 229.

SEQ ID NO: 229

MLNKYIEKRITDKITILNILLDIRSIELDELSTLTSLQSKSLLSILQELQETFEEELTFNLDTQQVQLIEHHSHQ
TNYYFHQLYNQSTILKILRFFLLQGNQSFNEFTQKEYISIATGYRVRQKCGLLLRSVGLDLVKNQVVGPEYRIRF
LIALLQFHFGIEIYDLNDGSMDWVTHMIVQSNSQLSHELLEITPDEYVHFSILVALTWKRREFPLEFPESKEFEK
LKNLFMYPILMEHCQTYLEPHANMTFTQEELDYIFLVYCSANSSFSKDKWNQEKKTHTIQLILQHTRGKHLLSKF
KNILGNDISNSLSFLTALTFLTRTFLFGLQNLVPYYNYYEHYGIESDKPLYHISKAIVQEWMTEQKIEGVIDQHR
LYLFSLYLTETIFSSLPAIPIFIILNNQADVNLIKSIILRNFTDKVASVTGYNILISPPPSEEHLTEPLIIITTK
EYLPYVKKQYPKGKHHFLTIALDLHVSQQRLIYQTIVDIRKEAFDKRVAMIAKKAHYLL

ORF3_23FTW is a cell wall surface protein. An example of an amino acid sequence of ORF3_23FTW is set forth in SEQ ID NO: 230.

SEQ ID NO: 230

MKKVRKIFQKAVAGLCCISQLTAFSSIVALAETPETSPAIGKVVIKETGEGGALLGDAVFELKNNTDGTTVSQRT
EAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQWTVEVEKNGRTTVQGEQVENREEALSDQYPQTGTYPDVQTP
YQIIKVDGSEKNGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTVYEQKDKSVPLDVVILLD
NSNSMSNIRNKNARRAERAGEATRSLIDKITSDPENRVALVTYASTIFDGTEFTVEKGVADKNGKRLNDSLFWNY
DQTSFTTNTKDYSYLKLTNDKNDIVELKNKVPTEAEDHDGNRLMYQFGATFTQKALMKADEILTQQARQNSQKVI
FHITDGVPTMSYPINFNHATFAPSYQNQLNAFFSKSPNKDGILLSDFITQATSGEHTIVRGDGQSYQMFTDKTVY
EKGAPAAFPVKPEKYSEMKAAGYAVIGDPINGGYIWLNWRESILAYPFNSNTAKITNHGDPTRWYYNGNIAPDGY
DVFTVGIGINGDPGTDEATATSFMQSISSKPENYTNVTDTTKILEQLNRYFHTIVTEKKSIENGTITDPMGELID
LQLGTDGRFDPADYTLTANDGSRLENGQAVGGPQNDGGLLKNAKVLYDTTEKRIRVTGLYLGTDEKVTLTYNVRL
NDEFVSNKFYDTNGRTTLHPKEVEQNTVRDFPIPKIRDVRKYPEITISKEKKLGDIEFIKVNKNDKKPLRDAVFS
LQKQHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVT
SIVPQDIPAGYEFTNDKHYITNEPIPPKREYPRTGGIGMLPFYLIGCMMMGGVLLYTRKHP

ORF4_23FTW is a cell wall surface protein. An example of an amino acid sequence of ORF4_23FTW is set forth in SEQ ID NO: 231.

SEQ ID NO: 231

MKSINKFLTILAALLLTVSSLFSAATVFAAEQKTKTLTVHKLLMTDQELDAWNSDAITTAGYDGSQNFEQFKQLQ
GVPQGVTEISGVAFELQSYTGPQGKEQENLTNDAVWTAVNKGVTTETGVKFDTEVLQGTYRLVEVRKESTYVGPN
GKVLTGMKAVPALITLPLVNQNGVVENAHVYPKNSEDKPTATKTFDTAAGFVDPGEKGLAIGTKVPYIVTTTIPK
NSTLATAFWSDEMTEGLDYNGDVVVNYNGQPLDNSHYTLEAGHNGFILKLNEKGLEAINGKDAEATITLKYTATL
NALAVADVPEANDVTFHYGNNPGHGNTPKPNKPKNGELTITKTWADAKDAPIAGVEVTFDLVNAQTGEVVKVPGH
ETGIVLNQTNNWTFTATGLDNNTEYKFVERTIKGYSADYQTITETGKIAVKNWKDENPEPINPEEPRVKTYGKKF
VKVDQKDERLKEAQFVVKNEQGKYLALKSAAQQAVNEKAAAEAKQALDAAIAAYTNAADKNAAQAVVDAAQKTYN
DNYRAARFGYVEVERKEDALVLTSNTDGQFQISGLAAGSYTLEETKAPEGFAKLGDVKFEVGAGSWNQGDFNYLK
DVQKNDATKVVNKKITIPQTGGIGTIIFAVAGAVIMGIAVYAYVKNNKDEDQLA

ORF5_23FTW is a cell wall surface protein. An example of an amino acid sequence of ORF5_23FTW is set forth in SEQ ID NO: 232.

SEQ ID NO: 232

MTMQKMQKMISRIFFVMALCFSLVWGAHAVQAQEDHTLVLQLENYQEVVSQLPSRDGHRLQVWKLDDSYSYDNRV
QIVRDLHSWDENKLSSFKKTSFEMTFLENQIEVSHIPNGLYYVRSIIQTDAVSYPAEFLFEMTDQTVEPLVIVAK
KADTVTTKVKLIKVDQDHNRLEGVGFKLVSVARDGSEKEVPLIGEYRYSSSGQVGRTLYTDKNGEIVVTNLPLGT
YRFKEVEPLAGYTVTTMDTDVQLVDHQLVTITVVNQKLPRGNVDFMKVDGRTNTSLQGAMFKVMKEENGHYTPVL
QNGKEVVVASGKDGRFRVEGLEYGTYYLWELQAPTGYVQLTSPVSFTIGKDTRKELVTVVKNNKRPRIDVPDTGE
ETLYILMLVAILLFGSGYYLTKKTNN

ORF6_23FTW is a putative sortase. An example of an amino acid sequence of ORF6_23FTW is set forth in SEQ ID NO: 233.

SEQ ID NO: 233

MLIKMVKTKKQKRNNLLLGVVFFIGMAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDS
LNNVVSGDPWSEEMKKKGRAEYARMLEIHERMGHVEIPVIDVDLPVYAGTAEEVLQQGAGQLEGTSLPIGGNSTH
AVITAHTGLPTAKMFTDLTKLKVGDKFYVHNIKEVMAYQVDQVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINT
HRLLVRGHRIPYVAEVEEEFIAANKLSHLYRYLFYVAVGLIVILLWIIRRLRKKKKQPEKALKALKAARKEVKVE
DGQQ

ORF7_23FTW is a putative sortase. An example of an amino acid sequence of ORF7_23FTW is set forth in SEQ ID NO: 234.

SEQ ID NO: 234

MDNSRRSRKKGTKKKKHPLILLLLIFLVGFAVAIYPLVSRYYYRIESNEVIKEFDETVSQMDKAELEERWRLAQAF
NATLKPSEILDPFTEQEKKKGVSEYANMLKVHERIGYVEIPAIDQEIPMYVGTSEEILQKGAGLLEGASLPVGGE
NTHTVVTAHRGLPTAELFSQLDKMKKGDVFYLHVLDQVLAYQVDQILTVEPNDFEPVLIQHGKDYATLLTCTPYM
INSHRLLVRGKRIPYTAPIAERNRAVRERGQFWLWLLLAALVMILVLSYGVYRHRRIVKGLEKQLEEHHVKG

ORF8_23FTW is a putative sortase. An example of an amino acid sequence of ORF8_23FTW is set forth in SEQ ID NO: 235.

SEQ ID NO: 235

MSKAKLQKLLGYLLMLVALVIPVYCFGQMVLQSLGQVKGHEIFSESVTADSYQEQLQRSLDYNQRLDSQNRIVDP
FLAEGYEVNYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLAMGLAHVDGTPLPVEGKGIRSVIAGHRAEPSH
VFFRHLDQLKVGDALYYDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAV
YQKSDPQTAAVARVAFTKEGQSVSRVATSQWLYRGLVVLAFLGILFVLWKLARLLRGK

As discussed above, a *S. pneumoniae* AI sequence is present in the 23F Poland 16 *S. pneumoniae* genome. Examples of *S. pneumoniae* AI sequences from 23F Poland 16 are set forth below.

ORF2_23FP is a transcriptional regulator. An example of an amino acid sequence of ORF2_23FP is set forth in SEQ ID NO: 236.

SEQ ID NO: 236

MLNKYIEKRITDKITILNILLDIRSIELDELSTLTSLQSKSLLSILQELQETFEEELTFNLDTQQVQLIEHHSHQ
TNYYFHQLYNQSTILKILRFFLLQGNQSFNEFTQKEYISIATGYRVRQKCGLLLRSVGLDLVKNQVVGPEYRIRF
LIALLQFHFGIEIYDLNDGSMDWVTHMIVQSNSQLSHELLEITPDEYVHFSILVALTWKRREFPLEFPESKEFEK
LKNLFMYPILMEHCQTYLEPHANMTFTQEELDYIFLVYCSANSSFSKDKWNQEKKTHTIQLILQHTRGKHLLSKF
KNILGNDISNSLSFLTALTFLTRTFLFGLQNLVPYYNYYEHYGIESDKPLYHISKAIVQEWMTEQKIEGVIDQHR
LYLFSLYLTETIFSSLPAIPIFIILNNQADVNLIKSIILRNFTDKVASVTGYNILISPPPSEEHLTEPLIIITTK
EYLPYVKKQYPKGKHHFLTIALDLHVSQQRLIYQTIVDIRKEAFDKRVAMIAKKAHYLL

ORF3_23FP is a cell wall surface protein. An example of an amino acid sequence of ORF3_23FP is set forth in SEQ ID NO: 237.

SEQ ID NO: 237

MKRVRKIEQKAVAGLCISQLIAFSSIVALAETPETSPAIGKVVIKETGEGGALLGDAVFELKNNTDGTTVSQRT
EAQTGEAIFSNIKPGTYTLTEAQPPVGYKPSTKQWTVEVEKNGRTTVQGEQVENREEALSDQYPQTGTYPDVQTP
YQIIKVDGSEKNGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDNQYGIELTVSGKTTVETKEASTPLDVVILLD
NSNSMSNIRHNHAHRAEKAGEATRALVDKITSNPDNRVALVTYGSTIFDGSEATVEKGVADANGKILNDSALWTF
DRTTFTAKTYNYSFLNLTSDPTDIQTIKDRIPSDAEELNKDKLMYQFGATFTQKALMTADDILTKQARPNSKKVI
FHITDGVPTMSYPINFKYTGTTQSYRTQLNNFKAKTPNSSGILLEDFVTWSADGEHKIVRGDGESYQMFTKKPVT
DQYGVHQILSITSMEQRAKLVSAGYRFYGTDLYLYWRDSILAYPFNSSTDWITNHGDPTTWYYNGNMAQDGYDVF
TVGVGVNGDPGTDEATATRFMQSISSSPDNYTNVADPSQILQELNRYFYTIVNEKKSIENGTITDPMGELIDFQL
GADGRFDPADYTLTANDGSSLVNNVPTGGPQNDGGLLKNAKVFYDTTEKRIRVTGLYLGTGEKVTLTYNVRLNDQ
FVSNKFYDTNGRTTLHPKEVEKNTVRDFPIPKIRDVRKYPEITIPKEKKLGEIEFIKINKNDKKPLRDAVFSLQK
QHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSEPAGYKPVQNKPIVAFQIVNGEVRDVTSIV
PQDIPAGYEFTNDKHYITNEPIPPKREYPRTGGIGMLPFYLIGCMMMGGVLLYTRKNP

ORF4_23FP is a cell wall surface protein. An example of an amino acid sequence of ORF4_23FP is set forth in SEQ ID NO: 238.

SEQ ID NO: 238

MKSINKFLTMLAALLLTASSLFSAATVFAADNVSTAPDAVTKTLTIHKLLLSEDDLKTWDTNGPKGYDGTQSSLK
DLTGVVAEEIPNVYFELQKYNLTDGKEKENLKDDSKWTTVHGGLTTKDGLKIETSTLKGVYRIREDRTKTTYVGP
NGQVLTGSKAVPALVTLPLVNNNGTVIDAHVFPKNSYNKPVVDKRIADTLNYNDQNGLSIGTKIPYVVNTTIPSN
ATFATSFWSDEMTEGLTYNEDVTITLNNVAMDQADYEVTKGINGFNLKLTEAGLAKINGKDADQKIQITYSATLN
SLAVADIPESNDITYHYGNHQDHGNTPKPTKPNNGQITVTKTWDSQPAPEGVKATVQLVNAKTGEKVGAPVELSE
NNWTYTWSGLDNSIEYKVEEEYNGYSAEYTVESKGKLGVKNWKDNNPAPINLEEPRVKTYGKKFVKVDQKDTRLE
NAQFVVKKADSNKYIAFKSTAQQAADEKAAATAKQKLDAAVAAYTNAADKQAAQALVDQAQQEYNVAYKEAKFGY
VEVAGKDEAMVLTSNTDGQFQISGLAAGTYKLEEIKAPEGFAKIDDVEFVVGAGSWNQGEFNYLKDVQKNDATKV
VNKKITIPQTGGIGTIIFAVAGAVIMGIAVYAYVKNNKDEDQLA

ORF5_23FP is a cell wall surface protein. An example of an amino acid sequence of ORF5_23FP is set forth in SEQ ID NO: 239.

SEQ ID NO: 239

MTMQKMQKMISRIFFVMALCFSLVWGAHAVQAQEDHTLVLQLENYQEVVSQLPSRDGHRLQVWKLDDSYSYDNRV
QIVRDLHSWDENKLSSFKKTSFEMTFLENQIEVSHIPNGLYYVRSIIQTDAVSYPAEFLFEMTDQTVEPLVIVAK
KADTVTTKVKLIKVDQDHNRLEGVGFKLVSVARDGSEKEVPLIGEYRYSSSGQVGRTLYTDKNGEIVVTNLPLGT
YRFKEVEPLAGYAVTTMDTDVQLVDHQLVTITVVNQKLPRGNVDFMKVDGRTNTSLQGAMFKVMKEENGHYTPVL
QNGKEVVVASGKDGRFRVEGLEYGTYYLWELQAPTGYVQLTSPVSFTIGKDTRKELVTVVKNNKRPRIDVPDTGE
ETLYILMLVAILLFGSGYYLTKKTNN

ORF6_23FP is a putative sortase. An example of an amino acid sequence of ORF6_23FP is set forth in SEQ ID NO: 240.

SEQ ID NO: 240

MLIKMAKTKKQKRNNLLLGVVFFIGIAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDS
LNNVVSGDPWSEEMKKKGRAEYARMLEIHERMGHVEIPAIDVDLPVYAGTAEEVLQQGAGHLEGTSLPIGGNSTH
AVITAHTGLPTAKMFTDLTKLKVGDKFYVHNIKEVMAYQVDQVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINT
HRLLVRGHRIPYVAEVEEEFIAANKLSHLYRYLFYVAVGLIVILLWIIRRLRKKKRQSERALKALKEATKEVKVE
DE

ORF7_23FP is a putative sortase. An example of an amino acid sequence of ORF7_23FP is set forth in SEQ ID NO: 241.

SEQ ID NO: 241

MSKSRYSRKKSVKKKKNPFILLLIFLVGLAVAMYPLVSRYYYRIESNEVIKEFDETVSQMDKAELEERWRLAQAF
NATLKPSEILDPFTEQEKKKGVSEYANMLKVHERIGYVEIPAIDQEIPMYVGTSEEILQKGAGLLEGASLPVGGE
NTHTVVTAHRGLPTAELFSQLDKMKKGDIFYLHVLDQVLAYQVDQIVTVEPNDFEPVLIQHGEDYATLLTCTPYM
INSHRLLVRGKRIPYTAPIAERNRAVRERGQFWLWLLLGAMAVILLLLYRVYRNRRIVKGLEKQLEGRHVKD

ORF8_23FP is a putative sortase. An example of an amino acid sequence of ORF8_23FP is set forth in SEQ ID NO: 242.

SEQ ID NO: 242 05/27239

MSRTKLRALLGYLLMLVACLIPIYCFGQMVLQSLGQVKGHATFVKSMTTEMYQEQQNHSLAYNQRLASQNRIVDP
FLAEGYEVNYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLGMGLAHVDGTPLPLDGTGIRSVIAGHRAEPSH
VFFRHLDQLKVGDALYYDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAV
YQKSDPQTAAVARVAFTKEGQSVSRVATSQWLYRGLVVLAFLGILFVLWKLARLLRGK

Immunogenic compositions of the invention comprising AI antigens may further comprise one or more antigenic agents. Preferred antigens include those listed below. Additionally, the compositions of the present invention may be used to treat or prevent infections caused by any of the below-listed microbes. Antigens for use in the immunogenic compositions include, but are not limited to, one or more of the following set forth below, or antigens derived from one or more of the following set forth below:

Bacterial Antigens

N. meningitides: a protein antigen from *N. meningitides* serogroup A, C, W135, Y, and/or B (1-7); an outer-membrane vesicle (OMV) preparation from *N. meningitides* serogroup B. (8, 9, 10, 11); a saccharide antigen, including LPS, from *N. meningitides* serogroup A, B, C W135 and/or Y, such as the oligosaccharide from serogroup C (see PCT/US99/09346; PCT IB98/01665; and PCT IB99/00103);

Streptococcus pneumoniae: a saccharide or protein antigen, particularly a saccharide from *Streptooccus pneumoniae*;

Streptococcus agalactiae: particularly, Group B streptococcus antigens;

Streptococcus pyogenes: particularly, Group A streptococcus antigens;

Enterococcus faecalis or Enterococcus faecium: Particularly a trisaccharide repeat or other *Enterococcus* derived antigens provided in US Patent No. 6,756,361;

Helicobacter pylori: including: Cag, Vac, Nap, HopX, HopY and/or urease antigen;

Bordetella pertussis: such as petussis holotoxin (PT) and filamentous haemagglutinin (FHA) from *B. pertussis*, optionally also combination with pertactin and/or agglutinogens 2 and 3 antigen;

Staphylococcus aureus: including *S. aureus* type 5 and 8 capsular polysaccharides optionally conjugated to nontoxic recombinant *Pseudomonas aeruginosa* exotoxin A, such as StaphVAX™, or antigens derived from surface proteins, invasins (leukocidin, kinases, hyaluronidase), surface factors that inhibit phagocytic engulfment (capsule, Protein A), carotenoids, catalase production, Protein A, coagulase, clotting factor, and/or membrane-damaging toxins (optionally detoxified) that lyse eukaryotic cell membranes (hemolysins, leukotoxin, leukocidin);

Staphylococcus epidermis: particularly, *S. epidermidis* slime-associated antigen (SAA);

Staphylococcus saprophyticus: (causing urinary tract infections) particularly the 160 kDa hemagglutinin of *S. saprophyticus* antigen;

Pseudomonas aeruginosa: particularly, endotoxin A, Wzz protein, *P. aeruginosa* LPS, more particularly LPS isolated from PAO1 (O5 serotype), and/or Outer Membrane Proteins, including Outer Membrane Proteins F (OprF) (*Infect Immun.* 2001 May; 69(5): 3510-3515);

Bacillus anthracis (anthrax): such as *B. anthracis* antigens (optionally detoxified) from A-components (lethal factor (LF) and edema factor (EF)), both of which can share a common B-component known as protective antigen (PA);

Moraxella catarrhalis: (respiratory) including outer membrane protein antigens (HMW-OMP), C-antigen, and/or LPS;

Yersinia pestis (plague): such as F1 capsular antigen (*Infect Immun.* 2003 Jan; 71(1)): 374-383, LPS (*Infect Immun.* 1999 Oct; 67(10): 5395), Yersinia pestis V antigen (*Infect Immun.* 1997 Nov; 65(11): 4476-4482);

Yersinia enterocolitica (gastrointestinal pathogen): particularly LPS (*Infect Immun.* 2002 August; 70(8): 4414);

Yersinia pseudotuberculosis: gastrointestinal pathogen antigens;

Mycobacterium tuberculosis: such as lipoproteins, LPS, BCG antigens, a fusion protein of antigen 85B (Ag85B) and/or ESAT-6 optionally formulated in cationic lipid vesicles (*Infect Immun.* 2004 October; 72(10): 6148), Mycobacterium tuberculosis (Mtb) isocitrate dehydrogenase associated antigens (*Proc Natl Acad Sci U S A.* 2004 Aug 24; 101(34): **12652), and/or MPT51 antigens** (*Infect Immun.* 2004 July; 72(7): 3829);

Legionella pneumophila (Legionnairs' Disease): L. pneumophila antigens -- optionally derived from cell lines with disrupted *asd* genes (*Infect Immun.* 1998 May; 66(5): 1898);

Rickettsia: including outer membrane proteins, including the outer membrane protein A and/or B (OmpB) (*Biochim Biophys Acta.* 2004 Nov 1;1702(2):145), LPS, and surface protein antigen (SPA) (*J Autoimmun.* 1989 Jun;2 Suppl:81);

E. coli: including antigens from enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAggEC), diffusely adhering *E. coli* (DAEC), enteropathogenic *E. coli* (EPEC), and/or enterohemorrhagic *E. coli* (EHEC);

Vibrio cholerae: including proteinase antigens, LPS, particularly lipopolysaccharides of Vibrio cholerae II, O1 Inaba O-specific polysaccharides, V. cholera O139, antigens of IEM108 vaccine (*Infect Immun.* 2003 Oct;71(10):5498-504), and/or Zonula occludens toxin (Zot);

Salmonella typhi (typhoid fever): including capsular polysaccharides preferably conjugates (Vi, i.e. vax-TyVi);

Salmonella typhimurium (gastroenteritis): antigens derived therefrom are contemplated for microbial and cancer therapies, including angiogenesis inhibition and modulation of flk;

Listeria monocytogenes (sytemic infections in immunocompromised or elderly people, infections of fetus): antigens derived from L. monocytogenes are preferably used as carriers/vectors for intracytoplasmic delivery of conjugates/associated compositions of the present invention;

Porphyromonas gingivalis: particularly, *P. gingivalis* outer membrane protein (OMP);

Tetanus: such as tetanus toxoid (TT) antigens, preferably used as a carrier protein in conjunction/conjugated with the compositions of the present invention;

Diphtheria: such as a diphtheria toxoid, preferably CRM_{197}, additionally antigens capable of modulating, inhibiting or associated with ADP ribosylation are contemplated for combination/co-administration/conjugation with the compositions of the present invention, the diphtheria toxoids are preferably used as carrier proteins;

Borrelia burgdorferi (Lyme disease): such as antigens associated with P39 and P13 (an integral membrane protein, *Infect Immun.* 2001 May; 69(5): 3323-3334), VlsE Antigenic Variation Protein (*J Clin Microbiol.* 1999 Dec; 37(12): 3997);

Haemophilus influenzae B: such as a saccharide antigen therefrom;

Klebsiella: such as an OMP, including OMP A, or a polysaccharide optionally conjugated to tetanus toxoid;

Neiserria gonorrhoeae: including, a Por (or porin) protein, such as PorB (*see* Zhu *et al.*, Vaccine (2004) 22:660 – 669), a transferring binding protein, such as TbpA and TbpB (See Price *et al.*, Infection and Immunity (2004) 71(1):277 – 283), a opacity protein (such as Opa), a reduction-modifiable protein (Rmp), and outer membrane vesicle (OMV) preparations (*see* Plante *et al.*, J Infectious Disease (2000) 182:848 – 855), also see *e.g.* WO99/24578, WO99/36544, WO99/57280, WO02/079243);

Chlamydia pneumoniae: particularly *C. pneumoniae* protein antigens;

Chlamydia trachomatis: including antigens derived from serotypes A, B, Ba and C are (agents of trachoma, a cause of blindness), serotypes L_1, L_2 & L_3 (associated with Lymphogranuloma venereum), and serotypes, D-K;

Treponema pallidum (Syphilis): particularly a TmpA antigen; and

Haemophilus ducreyi (causing chancroid): including outer membrane protein (DsrA).

Where not specifically referenced, further bacterial antigens of the invention may be capsular antigens, polysaccharide antigens or protein antigens of any of the above. Further bacterial antigens may also include an outer membrane vesicle (OMV) preparation. Additionally, antigens include live, attenuated, split, and/or purified versions of any of the aforementioned bacteria. The bacterial or microbial derived antigens of the present invention may be gram-negative or gram-positive and aerobic or anaerobic.

Additionally, any of the above bacterial-derived saccharides (polysaccharides, LPS, LOS or oligosaccharides) can be conjugated to another agent or antigen, such as a carrier protein (for example CRM_{197}). Such conjugation may be direct conjugation effected by reductive amination of carbonyl moieties on the saccharide to amino groups on the protein, as provided in US Patent No. 5,360,897 and *Can J Biochem Cell Biol.* 1984 May;62(5):270-5. Alternatively, the saccharides can be conjugated through a linker, such as, with succinamide or other linkages provided in *Bioconjugate Techniques,* 1996 and *CRC, Chemistry of Protein Conjugation and Cross-Linking*, 1993.

Viral Antigens

Influenza: including whole viral particles (attenuated), split, or subunit comprising hemagglutinin (HA) and/or neuraminidase (NA) surface proteins, the influenza antigens may be derived from chicken embryos or propogated on cell culture, and/or the influenza antigens are derived from influenza type A, B, and/or C, among others;

Respiratory syncytial virus (RSV): including the F protein of the A2 strain of RSV (*J Gen Virol.* 2004 Nov; 85(Pt 11):3229) and/or G glycoprotein;

Parainfluenza virus (PIV): including PIV type 1, 2, and 3, preferably containing hemagglutinin, neuraminidase and/or fusion glycoproteins;

Poliovirus: including antigens from a family of picornaviridae, preferably poliovirus antigens such as OPV or, preferably IPV;

Measles: including split measles virus (MV) antigen optionally combined with the Protollin and or antigens present in MMR vaccine;

Mumps: including antigens present in MMR vaccine;

Rubella: including antigens present in MMR vaccine as well as other antigens from Togaviridae, including dengue virus;

Rabies: such as lyophilized inactivated virus (RabAvert™);

Flaviviridae viruses: such as (and antigens derived therefrom) yelow fever virus, Japanese encephalitis virus, dengue virus (types 1, 2, 3, or 4), tick borne encephalitis virus, and West Nile virus;

Caliciviridae; antigens therefrom;

HIV: including HIV-1 or HIV-2 strain antigens, such as gag (p24gag and p55gag), env (gp160 and gp41), pol, tat, nef, rev vpu, miniproteins, (preferably p55 gag and gp140v delete) and antigens from the isolates HIV_{IIIb}, HIV_{SF2}, HIV_{LAV}, HIV_{LAI}, HIV_{MN}, HIV-1_{CM235}, HIV-1_{US4}, HIV-2; simian immunodeficiency virus (SIV) among others;

Rotavirus: including VP4, VP5, VP6, VP7, VP8 proteins (*Protein Expr Purif.* 2004 Dec;38(2):205) and/or NSP4;

Pestivirus: such as antigens from classical porcine fever virus, bovine viral diarrhoea virus, and/or border disease virus;

Parvovirus: such as parvovirus B19;

Coronavirus: including SARS virus antigens, particularly spike protein or proteases therefrom, as well as antigens included in WO 04/92360;

Hepatitis A virus: such as inactivated virus;

Hepatitis B virus: such as the surface and/or core antigens (sAg), as well as the presurface sequences, pre-S1 and pre-S2 (formerly called pre-S), as well as combinations of the above, such as sAg/pre-S1, sAg/pre-S2, sAg/pre-S1/pre-S2, and pre-S1/pre-S2, (see, e.g., AHBV Vaccines - *Human Vaccines and Vaccination*, pp. 159-176; and U.S. Patent Nos. 4,722,840, 5,098,704, 5,324,513;

Beames et al., *J. Virol.* (1995) 69:6833-6838, Birnbaum et al., *J. Virol.* (1990) 64:3319-3330; and Zhou et al., *J. Virol.* (1991) 65:5457-5464);

Hepatitis C virus: such as E1, E2, E1/E2 (see, Houghton et al., *Hepatology* (1991) 14:381), NS345 polyprotein, NS 345-core polyprotein, core, and/or peptides from the nonstructural regions (International Publication Nos. WO 89/04669; WO 90/11089; and WO 90/14436);

Delta hepatitis virus (HDV): antigens derived therefrom, particularly δ-antigen from HDV (see, e.g., U.S. Patent No. 5,378,814);

Hepatitis E virus (HEV); antigens derived therefrom;

Hepatitis G virus (HGV); antigens derived therefrom;

Varicella zoster virus: antigens derived from varicella zoster virus (VZV) (*J. Gen. Virol.* (1986) 67:1759);

Epstein-Barr virus: antigens derived from EBV (Baer et al., *Nature* (1984) 310:207);

Cytomegalovirus: CMV antigens, including gB and gH (*Cytomegaloviruses* (J.K. McDougall, ed., Springer-Verlag 1990) pp. 125-169);

Herpes simplex virus: including antigens from HSV-1 or HSV-2 strains and glycoproteins gB, gD and gH (McGeoch et al., *J. Gen. Virol.* (1988) 69:1531 and U.S. Patent No. 5,171,568);

Human Herpes Virus: antigens derived from other human herpesviruses such as HHV6 and HHV7; and

HPV: including antigens associated with or derived from human papillomavirus (HPV), for example, one or more of E1 – E7, L1, L2, and fusions thereof, particularly the compositions of the invention may include a virus-like particle (VLP) comprising the L1 major capsid protein, more particular still, the HPV antigens are protective against one or more of HPV serotypes 6, 11, 16 and/or 18.

Further provided are antigens, compostions, methods, and microbes included in *Vaccines*, 4th Edition (Plotkin and Orenstein ed. 2004); *Medical Microbiology* 4th Edition (Murray et al. ed. 2002); *Virology*, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991), which are contemplated in conjunction with the compositions of the present invention.

Additionally, antigens include live, attenuated, split, and/or purified versions of any of the aforementioned viruses.

Fungal Antigens

Fungal antigens for use herein, associated with vaccines include those described in: U.S. Pat. Nos. 4,229,434 and 4,368,191 for prophylaxis and treatment of trichopytosis caused by Trichophyton mentagrophytes; U.S. Pat. Nos. 5,277,904 and 5,284,652 for a broad spectrum dermatophyte vaccine for the prophylaxis of dermatophyte infection in animals, such as guinea pigs, cats, rabbits, horses and lambs, these antigens comprises a suspension of killed *T. equinum,* T. mentagrophytes (var. granulare), *M. canis* and/or *M. gypseum* in an effective amount optionally combined with an adjuvant;

U.S. Pat. Nos. 5,453,273 and 6,132,735 for a ringworm vaccine comprising an effective amount of a homogenized, formaldehyde-killed fungi, i.e., *Microsporum canis* culture in a carrier; U.S. Pat. No. 5,948,413 involving extracellular and intracellular proteins for pythiosis. Additional antigens identified within antifungal vaccines include Ringvac bovis LTF-130 and Bioveta.

Further, fungal antigens for use herein may be derived from Dermatophytres, including: *Epidermophyton floccusum, Microsporum audouini, Microsporum canis, Microsporum distortum, Microsporum equinum, Microsporum gypsum, Microsporum nanum, Trichophyton concentricum, Trichophyton equinum, Trichophyton gallinae, Trichophyton gypseum, Trichophyton megnini, Trichophyton mentagrophytes, Trichophyton quinckeanum, Trichophyton rubrum, Trichophyton schoenleini, Trichophyton tonsurans, Trichophyton verrucosum, T. verrucosum* var. album, var. discoides, var. ochraceum, *Trichophyton violaceum,* and/or *Trichophyton faviforme.*

Fungal pathogens for use as antigens or in derivation of antigens in conjunction with the compositions of the present invention comprise *Aspergillus fumigatus, Aspergillus flavus, Aspergillus niger, Aspergillus nidulans, Aspergillus terreus, Aspergillus sydowi, Aspergillus flavatus, Aspergillus glaucus, Blastoschizomyces capitatus, Candida albicans, Candida enolase, Candida tropicalis, Candida glabrata, Candida krusei, Candida parapsilosis, Candida stellatoidea, Candida kusei, Candida parakwsei, Candida lusitaniae, Candida pseudotropicalis, Candida guilliermondi, Cladosporium carrionii, Coccidioides immitis, Blastomyces dermatidis, Cryptococcus neoformans, Geotrichum clavatum, Histoplasma capsulatum, Klebsiella pneumoniae, Paracoccidioides brasiliensis, Pneumocystis carinii, Pythiumn insidiosum, Pityrosporum ovale, Sacharomyces cerevisae, Saccharomyces boulardii, Saccharomyces pombe, Scedosporium apiosperum, Sporothrix schenckii, Trichosporon beigelii, Toxoplasma gondii, Penicillium marneffei,* Malassezia spp., Fonsecaea spp., Wangiella spp., Sporothrix spp., Basidiobolus spp., Conidiobolus spp., Rhizopus spp, Mucor spp, Absidia spp, Mortierella spp, Cunninghamella spp, and Saksenaea spp.

Other fungi from which antigens are derived include Alternaria spp, Curvularia spp, Helminthosporium spp, Fusarium spp, Aspergillus spp, Penicillium spp, Monolinia spp, Rhizoctonia spp, Paecilomyces spp, Pithomyces spp, and Cladosporium spp.

Processes for producing a fungal antigens are well known in the art (see US Patent No. 6,333,164). In a preferred method a solubilized fraction extracted and separated from an insoluble fraction obtainable from fungal cells of which cell wall has been substantially removed or at least partially removed, characterized in that the process comprises the steps of: obtaining living fungal cells; obtaining fungal cells of which cell wall has been substantially removed or at least partially removed; bursting the fungal cells of which cell wall has been substantially removed or at least partially removed; obtaining an insoluble fraction; and extracting and separating a solubilized fraction from the insoluble fraction.

PCT/US2005/027239

STD Antigens

In particular embodiments, microbes (bacteria, viruses and/or fungi) against which the present compositions and methods can be implement include those that cause sexually transmitted diseases (STDs) and/or those that display on their surface an antigen that can be the target or antigen composition of the invention. In a preferred embodiment of the invention, compositions are combined with antigens derived from a viral or bacterial STD. Antigens derived from bacteria or viruses can be administered in conjunction with the compositions of the present invention to provide protection against at least one of the following STDs, among others: chlamydia, genital herpes, hepatitis (particularly HCV), genital warts, gonorrhoea, syphilis and/or chancroid (See, WO00/15255).

In another embodiment the compositions of the present invention are co-administered with an antigen for the prevention or treatment of an STD.

Antigens derived from the following viruses associated with STDs, which are described in greater detail above, are preferred for co-administration with the compositions of the present invention: hepatitis (particularly HCV), HPV, HIV, or HSV.

Additionally, antigens derived from the following bacteria associated with STDs, which are described in greater detail above, are preferred for co-administration with the compositions of the present invention: *Neiserria gonorrhoeae, Chlamydia pneumoniae, Chlamydia trachomatis, Treponema pallidum*, or *Haemophilus ducreyi.*

Respiratory Antigens

The antigen may be a respiratory antigen and could further be used in an immunogenic composition for methods of preventing and/or treating infection by a respiratory pathogen, including a virus, bacteria, or fungi such as respiratory syncytial virus (RSV), PIV, SARS virus, influenza, *Bacillus anthracis,* particularly by reducing or preventing infection and/or one or more symptoms of respiratory virus infection. A composition comprising an antigen described herein, such as one derived from a respiratory virus, bacteria or fungus is administered in conjunction with the compositions of the present invention to an individual which is at risk of being exposed to that particular respiratory microbe, has been exposed to a respiratory microbe or is infected with a respiratory virus, bacteria or fungus. The composition(s) of the present invention is/are preferably co-administered at the same time or in the same formulation with an antigen of the respiratory pathogen. Administration of the composition results in reduced incidence and/or severity of one or more symptoms of respiratory infection.

Pediatric/Geriatric Antigens

In one embodiment the compositions of the present invention are used in conjunction with an antigen for treatment of a pediatric population, as in a pediatric antigen. In a more particular embodiment the pediatric population is less than about 3 years old, or less than about 2 years, or less than about 1 years old. In another embodiment the pediatric antigen (in conjunction with the composition of the present invention) is administered multiple times over at least 1, 2, or 3 years.

In another embodiment the compositions of the present invention are used in conjunction with an antigen for treatment of a geriatric population, as in a geriatric antigen.

Other Antigens

Other antigens for use in conjunction with the compositions of the present include hospital acquired (nosocomial) associated antigens.

In another embodiment, parasitic antigens are contemplated in conjunction with the compositions of the present invention. Examples of parasitic antigens include those derived from organisms causing malaria and/or Lyme disease.

In another embodiment, the antigens in conjunction with the compositions of the present invention are associated with or effective against a mosquito born illness. In another embodiment, the antigens in conjunction with the compositions of the present invention are associated with or effective against encephalitis. In another embodiment the antigens in conjunction with the compositions of the present invention are associated with or effective against an infection of the nervous system.

In another embodiment, the antigens in conjunction with the compositions of the present invention are antigens transmissible through blood or body fluids.

Antigen Formulations

In other aspects of the invention, methods of producing microparticles having adsorbed antigens are provided. The methods comprise: (a) providing an emulsion by dispersing a mixture comprising (i) water, (ii) a detergent, (iii) an organic solvent, and (iv) a biodegradable polymer selected from the group consisting of a poly(α-hydroxy acid), a polyhydroxy butyric acid, a polycaprolactone, a polyorthoester, a polyanhydride, and a polycyanoacrylate. The polymer is typically present in the mixture at a concentration of about 1% to about 30% relative to the organic solvent, while the detergent is typically present in the mixture at a weight-to-weight detergent-to-polymer ratio of from about 0.00001:1 to about 0.1:1 (more typically about 0.0001:1 to about 0.1:1, about 0.001:1 to about 0.1:1, or about 0.005:1 to about 0.1:1); (b) removing the organic solvent from the emulsion; and (c) adsorbing an antigen on the surface of the microparticles. In certain embodiments, the biodegradable polymer is present at a concentration of about 3% to about 10% relative to the organic solvent.

Microparticles for use herein will be formed from materials that are sterilizable, non-toxic and biodegradable. Such materials include, without limitation, poly(α-hydroxy acid), polyhydroxybutyric acid, polycaprolactone, polyorthoester, polyanhydride, PACA, and polycyanoacrylate. Preferably, microparticles for use with the present invention are derived from a poly(α-hydroxy acid), in particular, from a poly(lactide) ("PLA") or a copolymer of D,L-lactide and glycolide or glycolic acid, such as a poly(D,L-lactide-co-glycolide) ("PLG" or "PLGA"), or a copolymer of D,L-lactide and caprolactone. The microparticles may be derived from any of various polymeric starting materials which have a variety of molecular weights and, in the case of the copolymers such as PLG, a variety of lactide:glycolide ratios, the selection of which will be largely a

matter of choice, depending in part on the coadministered macromolecule. These parameters are discussed more fully below.

Further antigens may also include an outer membrane vesicle (OMV) preparation.

Additional formulation methods and antigens (especially tumor antigens) are provided in U.S. Patent Serial No. 09/581,772.

Antigen References

The following references include antigens useful in conjunction with the compositions of the present invention:

1 International patent application WO99/24578
2 International patent application WO99/36544.
3 International patent application WO99/57280.
4 International patent application WO00/22430.
5 Tettelin et al. (2000) Science 287:1809-1815.
6 International patent application WO96/29412.
7 Pizza et al. (2000) Science 287:1816-1820.
8 PCT WO 01/52885.
9 Bjune et al. (1991) Lancet 338(8775).
10 Fuskasawa et al. (1999) Vaccine 17:2951-2958.
11 Rosenqist et al. (1998) Dev. Biol. Strand 92:323-333.
12 Constantino et al. (1992) Vaccine 10:691-698.
13 Constantino et al. (1999) Vaccine 17:1251-1263.
14 Watson (2000) Pediatr Infect Dis J 19:331-332.
15 Rubin (20000) Pediatr Clin North Am 47:269-285,v.
16 Jedrzejas (2001) Microbiol Mol Biol Rev 65:187-207.
17 International patent application filed on 3rd July 2001 claiming priority from GB-0016363.4;WO 02/02606; PCT IB/01/00166.
18 Kalman et al. (1999) Nature Genetics 21:385-389.
19 Read et al. (2000) Nucleic Acids Res 28:1397-406.
20 Shirai et al. (2000) J. Infect. Dis 181(Suppl 3):S524-S527.
21 International patent application WO99/27105.
22 International patent application WO00/27994.
23 International patent application WO00/37494.
24 International patent application WO99/28475.
25 Bell (2000) Pediatr Infect Dis J 19:1187-1188.
26 Iwarson (1995) APMIS 103:321-326.
27 Gerlich et al. (1990) Vaccine 8 Suppl:S63-68 & 79-80.
28 Hsu et al. (1999) Clin Liver Dis 3:901-915.
29 Gastofsson et al. (1996) N. Engl. J. Med. 334-:349-355.
30 Rappuoli et al. (1991) TIBTECH 9:232-238.
31 Vaccines (1988) eds. Plotkin & Mortimer. ISBN 0-7216-1946-0.
32 Del Guidice et al. (1998) Molecular Aspects of Medicine 19:1-70.
33 International patent application WO93/018150.
34 International patent application WO99/53310.
35 International patent application WO98/04702.
36 Ross et al. (2001) Vaccine 19:135-142.
37 Sutter et al. (2000) Pediatr Clin North Am 47:287-308.
38 Zimmerman & Spann (1999) Am Fan Physician 59:113-118, 125-126.
39 Dreensen (1997) Vaccine 15 Suppl"S2-6.
40 MMWR Morb Mortal Wkly rep 1998 Jan 16:47(1):12, 9.
41 McMichael (2000) Vaccine19 Suppl 1:S101-107.

42 Schuchat (1999) Lancet 353(9146):51-6.
43 GB patent applications 0026333.5, 0028727.6 & 0105640.7.
44 Dale (1999) Infect Disclin North Am 13:227-43, viii.
45 Ferretti et al. (2001) PNAS USA 98: 4658-4663.
46 Kuroda et al. (2001) Lancet 357(9264):1225-1240; see also pages 1218-1219.
47 Ramsay et al. (2001) Lancet 357(9251):195-196.
48 Lindberg (1999) Vaccine 17 Suppl.2:S28-36.
49 Buttery & Moxon (2000) J R Coil Physicians Long 34:163-168.
50 Ahmad & Chapnick (1999) Infect Dis Clin North Am 13:113-133, vii.
51 Goldblatt (1998) J. Med. Microbiol. 47:663-567.
52 European patent 0 477 508.
53 U.S. Patent No. 5,306,492.
54 International patent application WO98/42721.
55 Conjugate Vaccines (eds. Cruse et al.) ISBN 3805549326, particularly vol. 10:48-114.
56 Hermanson (1996) Bioconjugate Techniques ISBN: 012323368 & 012342335X.
57 European patent application 0372501.
58 European patent application 0378881.
59 European patent application 0427347.
60 International patent application WO93/17712.
61 International patent application WO98/58668.
62 European patent application 0471177.
63 International patent application WO00/56360.
64 International patent application WO00/67161.

The contents of all of the above cited patents, patent applications and journal articles are incorporated by reference as if set forth fully herein.

There may be an upper limit to the number of Gram positive bacterial proteins which will be in the compositions of the invention. Preferably, the number of Gram positive bacterial proteins in a composition of the invention is less than 20, less than 19, less than 18, less than 17, less than 16, less than 15, less than 14, less than 13, less than 12, less than 11, less than 10, less than 9, less than 8, less than 7, less than 6, less than 5, less than 4, or less than 3. Still more preferably, the number of Gram positive bacterial proteins in a composition of the invention is less than 6, less than 5, or less than 4. Still more preferably, the number of Gram positive bacterial proteins in a composition of the invention is 3.

The Gram positive bacterial proteins and polynucleotides used in the invention are preferably isolated, *i.e.*, separate and discrete, from the whole organism with which the molecule is found in nature or, when the polynucleotide or polypeptide is not found in nature, is sufficiently free of other biological macromolecules so that the polynucleotide or polypeptide can be used for its intended purpose.

Fusion Proteins: GBS AI sequences

The GBS AI proteins used in the invention may be present in the composition as individual separate polypeptides, but it is preferred that at least two (*i.e.* 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18) of the antigens are expressed as a single polypeptide chain (a "hybrid" or "fusion" polypeptide). Such fusion polypeptides offer two principal advantages: first, a polypeptide that may be unstable or poorly expressed on its own can be assisted by adding a suitable fusion partner that

overcomes the problems: second, commercial manufacture is simplified as only one expression and purification need be employed in order to produce two polypeptides which are both antigenically useful.

The fusion polypeptide may comprise one or more AI polypeptide sequences. Preferably, the fusion comprises an AI surface protein sequence. Preferably, the fusion polypeptide includes one or more of GBS 80, GBS 104, and GBS 67. Most preferably, the fusion peptide includes a polypeptide sequence from GBS 80. Accordingly, the invention includes a fusion peptide comprising a first amino acid sequence and a second amino acid sequence, wherein said first and second amino acid sequences are selected from a GBS AI surface protein or a fragment thereof. Preferably, the first and second amino acid sequences in the fusion polypeptide comprise different epitopes.

Hybrids (or fusions) consisting of amino acid sequences from two, three, four, five, six, seven, eight, nine, or ten GBS antigens are preferred. In particular, hybrids consisting of amino acid sequences from two, three, four, or five GBS antigens are preferred.

Different hybrid polypeptides may be mixed together in a single formulation. Within such combinations, a GBS antigen may be present in more than one hybrid polypeptide and/or as a non-hybrid polypeptide. It is preferred, however, that an antigen is present either as a hybrid or as a non-hybrid, but not as both.

Hybrid polypeptides can be represented by the formula **NH_2-A-{-X-L-}$_n$-B-COOH**, wherein: X is an amino acid sequence of a GBS AI protein or a fragment thereof; L is an optional linker amino acid sequence; A is an optional N-terminal amino acid sequence; B is an optional C-terminal amino acid sequence; and *n* is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15.

If a -X- moiety has a leader peptide sequence in its wild-type form, this may be included or omitted in the hybrid protein. In some embodiments, the leader peptides will be deleted except for that of the -X- moiety located at the N-terminus of the hybrid protein *i.e.* the leader peptide of X_1 will be retained, but the leader peptides of X_2 ... X_n will be omitted. This is equivalent to deleting all leader peptides and using the leader peptide of X_1 as moiety -A-.

For each *n* instances of {-X-L-}, linker amino acid sequence -L- may be present or absent. For instance, when *n*=2 the hybrid may be NH_2-X_1-L_1-X_2-L_2-COOH, NH_2-X_1-X_2-COOH, NH_2-X_1-L_1-X_2-COOH, NH_2-X_1-X_2-L_2-COOH, *etc.* Linker amino acid sequence(s) -L- will typically be short (*e.g.* 20 or fewer amino acids *i.e.* 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples comprise short peptide sequences which facilitate cloning, poly-glycine linkers (*i.e.* comprising Gly_n where $n = 2, 3, 4, 5, 6, 7, 8, 9, 10$ or more), and histidine tags (*i.e.* His_n where $n = 3, 4, 5, 6, 7, 8, 9, 10$ or more). Other suitable linker amino acid sequences will be apparent to those skilled in the art. A useful linker is GSGGGG, with the Gly-Ser dipeptide being formed from a *Bam*HI restriction site, thus aiding cloning and manipulation, and the $(Gly)_4$ tetrapeptide being a typical poly-glycine linker.

-A- is an optional N-terminal amino acid sequence. This will typically be short (*e.g.* 40 or fewer amino acids *i.e.* 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19,

18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include leader sequences to direct protein trafficking, or short peptide sequences which facilitate cloning or purification (*e.g.* histidine tags *i.e.* His_n where $n = 3, 4, 5, 6, 7, 8, 9, 10$ or more). Other suitable N-terminal amino acid sequences will be apparent to those skilled in the art. If X_1 lacks its own N-terminus methionine, -A- is preferably an oligopeptide (*e.g.* with 1, 2, 3, 4, 5, 6, 7 or 8 amino acids) which provides a N-terminus methionine.

-B- is an optional C-terminal amino acid sequence. This will typically be short (*e.g.* 40 or fewer amino acids *i.e.* 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include sequences to direct protein trafficking, short peptide sequences which facilitate cloning or purification (*e.g.* comprising histidine tags *i.e.* His_n where $n = 3, 4, 5, 6, 7, 8, 9, 10$ or more), or sequences which enhance protein stability. Other suitable C-terminal amino acid sequences will be apparent to those skilled in the art.

Most preferably, n is 2 or 3.

Fusion Proteins: Gram positive bacteria AI sequences

The Gram positive bacteria AI proteins used in the invention may be present in the composition as individual separate polypeptides, but it is preferred that at least two (*i.e.* 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18) of the antigens are expressed as a single polypeptide chain (a "hybrid" or "fusion" polypeptide). Such fusion polypeptides offer two principal advantages: first, a polypeptide that may be unstable or poorly expressed on its own can be assisted by adding a suitable fusion partner that overcomes the problem; second, commercial manufacture is simplified as only one expression and purification need be employed in order to produce two polypeptides which are both antigenically useful.

The fusion polypeptide may comprise one or more AI polypeptide sequences. Preferably, the fusion comprises an AI surface protein sequence. Accordingly, the invention includes a fusion peptide comprising a first amino acid sequence and a second amino acid sequence, wherein said first and second amino acid sequences are selected from a Gram positive bacteria AI protein or a fragment thereof. Preferably, the first and second amino acid sequences in the fusion polypeptide comprise different epitopes.

Hybrids (or fusions) consisting of amino acid sequences from two, three, four, five, six, seven, eight, nine, or ten Gram positive bacteria antigens are preferred. In particular, hybrids consisting of amino acid sequences from two, three, four, or five Gram positive bacteria antigens are preferred.

Different hybrid polypeptides may be mixed together in a single formulation. Within such combinations, a Gram positive bacteria AI sequence may be present in more than one hybrid polypeptide and/or as a non-hybrid polypeptide. It is preferred, however, that an antigen is present either as a hybrid or as a non-hybrid, but not as both.

Hybrid polypeptides can be represented by the formula $\mathbf{NH_2\text{-}A\text{-}\{\text{-}X\text{-}L\text{-}\}_n\text{-}B\text{-}COOH}$, wherein: X is an amino acid sequence of a Gram positive bacteria AI sequence or a fragment thereof; L is an

optional linker amino acid sequence; A is an optional N-terminal amino acid sequence; B is an optional C-terminal amino acid sequence; and *n* is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15.

If a -X- moiety has a leader peptide sequence in its wild-type form, this may be included or omitted in the hybrid protein. In some embodiments, the leader peptides will be deleted except for that of the -X- moiety located at the N-terminus of the hybrid protein *i.e.* the leader peptide of X_1 will be retained, but the leader peptides of X_2 ... X_n will be omitted. This is equivalent to deleting all leader peptides and using the leader peptide of X_1 as moiety -A-.

For each *n* instances of {-X-L-}, linker amino acid sequence -L- may be present or absent. For instance, when *n*=2 the hybrid may be NH_2-X_1-L_1-X_2-L_2-COOH, NH_2-X_1-X_2-COOH, NH_2-X_1-L_1-X_2-COOH, NH_2-X_1-X_2-L_2-COOH, *etc.* Linker amino acid sequence(s) -L- will typically be short (*e.g.* 20 or fewer amino acids *i.e.* 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples comprise short peptide sequences which facilitate cloning, poly-glycine linkers (*i.e.* comprising Gly_n where $n = 2, 3, 4, 5, 6, 7, 8, 9, 10$ or more), and histidine tags (*i.e.* His_n where $n = 3, 4, 5, 6, 7, 8, 9, 10$ or more). Other suitable linker amino acid sequences will be apparent to those skilled in the art. A useful linker is GSGGGG, with the Gly-Ser dipeptide being formed from a *Bam*HI restriction site, thus aiding cloning and manipulation, and the $(Gly)_4$ tetrapeptide being a typical poly-glycine linker.

-A- is an optional N-terminal amino acid sequence. This will typically be short (*e.g.* 40 or fewer amino acids *i.e.* 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include leader sequences to direct protein trafficking, or short peptide sequences which facilitate cloning or purification (*e.g.* histidine tags *i.e.* His_n where $n = 3, 4, 5, 6, 7, 8, 9, 10$ or more). Other suitable N-terminal amino acid sequences will be apparent to those skilled in the art. If X_1 lacks its own N-terminus methionine, -A- is preferably an oligopeptide (*e.g.* with 1, 2, 3, 4, 5, 6, 7 or 8 amino acids) which provides a N-terminus methionine.

-B- is an optional C-terminal amino acid sequence. This will typically be short (*e.g.* 40 or fewer amino acids *i.e.* 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include sequences to direct protein trafficking, short peptide sequences which facilitate cloning or purification (*e.g.* comprising histidine tags *i.e.* His_n where $n = 3, 4, 5, 6, 7, 8, 9, 10$ or more), or sequences which enhance protein stability. Other suitable C-terminal amino acid sequences will be apparent to those skilled in the art.

Most preferably, *n* is 2 or 3.

Antibodies: GBS AI sequences

The GBS AI proteins of the invention may also be used to prepare antibodies specific to the GBS AI proteins. The antibodies are preferably specific to the an oligomeric or hyper-oligomeric form of an AI protein. The invention also includes combinations of antibodies specific to GBS AI proteins selected to provide protection against an increased range of GBS serotypes and strain isolates. For example, a combination may comprise a first and second antibody, wherein said first

antibody is specific to a first GBS AI protein and said second antibody is specific to a second GBS AI protein. Preferably, the nucleic acid sequence encoding said first GBS AI protein is not present in a GBS genome comprising a polynucleotide sequence encoding for said second GBS AI protein. Preferably, the nucleic acid sequence encoding said first and second GBS AI proteins are present in the genomes of multiple GBS serotypes and strain isolates.

The GBS specific antibodies of the invention include one or more biological moieties that, through chemical or physical means, can bind to or associate with an epitope of a GBS polypeptide. The antibodies of the invention include antibodies which specifically bind to a GBS AI protein. The invention includes antibodies obtained from both polyclonal and monoclonal preparations, as well as the following: hybrid (chimeric) antibody molecules (see, for example, Winter *et al.* (1991) *Nature* 349: 293-299; and US Patent No. 4,816,567; F(ab')$_2$ and F(ab) fragments; F_v molecules (non-covalent heterodimers, see, for example, Inbar *et al.* (1972) *Proc Natl Acad Sci USA* 69:2659-2662; and Ehrlich *et al.* (1980) *Biochem* 19:4091-4096); single-chain Fv molecules (sFv) (see, for example, Huston *et al.* (1988) *Proc Natl Acad Sci USA* 85:5897-5883); dimeric and trimeric antibody fragment constructs; minibodies (see, *e.g.*, Pack *et al.* (1992) *Biochem* 31:1579-1584; Cumber *et al.* (1992) *J Immunology* 149B: 120-126); humanized antibody molecules (see, for example, Riechmann *et al.* (1988) *Nature* 332:323-327; Verhoeyan *et al.* (1988) *Science* 239:1534-1536; and U.K. Patent Publication No. GB 2,276,169, published 21 September 1994); and, any functional fragments obtained from such molecules, wherein such fragments retain immunological binding properties of the parent antibody molecule. The invention further includes antibodies obtained through non-conventional processes, such as phage display.

Preferably, the GBS specific antibodies of the invention are monoclonal antibodies. Monoclonal antibodies of the invention include an antibody composition having a homogeneous antibody population. Monoclonal antibodies of the invention may be obtained from murine hybridomas, as well as human monoclonal antibodies obtained using human rather than murine hybridomas. See, *e.g.*, Cote, *et al. Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, p 77.

The antibodies of the invention may be used in diagnostic applications, for example, to detect the presence or absence of GBS in a biological sample. The antibodies of the invention may also be used in the prophylactic or therapeutic treatment of GBS infection.

Antibodies: Gram positive bacteria AI sequences

The Gram positive bacteria AI proteins of the invention may also be used to prepare antibodies specific to the Gram positive bacteria AI proteins. The antibodies are preferably specific to the an oligomeric or hyper-oligomeric form of an AI protein. The invention also includes combinations of antibodies specific to Gram positive bacteria AI proteins selected to provide protection against an increased range of Gram positive bacteria genus, species, serotypes and strain isolates.

For example, a combination may comprise a first and second antibody, wherein said first antibody is specific to a first Gram positive bacteria AI protein and said second antibody is specific to a second Gram positive bacteria AI protein. Preferably, the nucleic acid sequence encoding said first Gram positive bacteria AI protein is not present in a Gram positive bacterial genome comprising a polynucleotide sequence encoding for said second Gram positive bacteria AI protein. Preferably, the nucleic acid sequence encoding said first and second Gram positive bacteria AI proteins are present in the genomes of multiple Gram positive bacteria genus, species, serotypes or strain isolates.

As an example of an instance where the combination of antibodies provides protection against an increased range of bacteria serotypes, the first antibody may be specific to a first GAS AI protein and the second antibody may be specific to a second GAS AI protein. The first GAS AI protein may comprise a GAS AI-1 surface protein, while the second GAS AI protein may comprise a GAS AI-2 or AI-3 surface protein.

As an example of an instance where the combination of antibodies provides protection against an increased range of bacterial species, the first antibody may be specific to a GBS AI protein and the second antibody may be specific to a GAS AI protein. Alternatively, the first antibody may be specific to a GAS AI protein and the second antibody may be specific to a *S. pneumoniae* AI protein.

The Gram positive specific antibodies of the invention include one or more biological moieties that, through chemical or physical means, can bind to or associate with an epitope of a Gram positive bacteria AI polypeptide. The antibodies of the invention include antibodies which specifically bind to a Gram positive bacteria AI protein. The invention includes antibodies obtained from both polyclonal and monoclonal preparations, as well as the following: hybrid (chimeric) antibody molecules (see, for example, Winter *et al.* (1991) *Nature* 349: 293-299; and US Patent No. 4,816,567; $F(ab')_2$ and F(ab) fragments; F_v molecules (non-covalent heterodimers, see, for example, Inbar *et al.* (1972) *Proc Natl Acad Sci USA* 69:2659-2662; and Ehrlich *et al.* (1980) *Biochem* 19:4091-4096); single-chain Fv molecules (sFv) (see, for example, Huston *et al.* (1988) *Proc Natl Acad Sci USA* 85:5897-5883); dimeric and trimeric antibody fragment constructs; minibodies (see, *e.g.*, Pack *et al.* (1992) *Biochem* 31:1579-1584; Cumber *et al.* (1992) *J Immunology* 149B: 120-126); humanized antibody molecules (see, for example, Riechmann *et al.* (1988) *Nature* 332:323-327; Verhoeyan *et al.* (1988) *Science* 239:1534-1536; and U.K. Patent Publication No. GB 2,276,169, published 21 September 1994); and, any functional fragments obtained from such molecules, wherein such fragments retain immunological binding properties of the parent antibody molecule. The invention further includes antibodies obtained through non-conventional processes, such as phage display.

Preferably, the Gram positive specific antibodies of the invention are monoclonal antibodies. Monoclonal antibodies of the invention include an antibody composition having a homogeneous antibody population. Monoclonal antibodies of the invention may be obtained from murine hybridomas, as well as human monoclonal antibodies obtained using human rather than murine

PCT/US2005/027239
hybridomas. See *e.g.* Cote, *et al. Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, p 77.

The antibodies of the invention may be used in diagnostic applications, for example, to detect the presence or absence of Gram positive bacteria in a biological sample. The antibodies of the invention may also be used in the prophylactic or therapeutic treatment of Gram positive bacteria infection.

Nucleic Acids

The invention provides nucleic acids encoding the Gram positive bacteria sequences and/or the hybrid fusion polypeptides of the invention. The invention also provides nucleic acid encoding the GBS antigens and/or the hybrid fusion polypeptides of the invention. Furthermore, the invention provides nucleic acid which can hybridise to these nucleic acids, preferably under "high stringency" conditions (*e.g.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Polypeptides of the invention can be prepared by various means (*e.g.* recombinant expression, purification from cell culture, chemical synthesis, *etc.*) and in various forms (*e.g.* native, fusions, non-glycosylated, lipidated, *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other GAS or host cell proteins).

Nucleic acid according to the invention can be prepared in many ways (*e.g.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself, *etc.*) and can take various forms (*e.g.* single stranded, double stranded, vectors, probes, *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other GBS or host cell nucleic acids).

The term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones (*e.g.* phosphorothioates, *etc.*), and also peptide nucleic acids (PNA), *etc.* The invention includes nucleic acid comprising sequences complementary to those described above (*e.g.* for antisense or probing purposes).

The invention also provides a process for producing a polypeptide of the invention, comprising the step of culturing a host cell transformed with nucleic acid of the invention under conditions which induce polypeptide expression.

The invention provides a process for producing a polypeptide of the invention, comprising the step of synthesising at least part of the polypeptide by chemical means.

The invention provides a process for producing nucleic acid of the invention, comprising the step of amplifying nucleic acid using a primer-based amplification method (*e.g.* PCR).

The invention provides a process for producing nucleic acid of the invention, comprising the step of synthesising at least part of the nucleic acid by chemical means.

Purification and Recombinant Expression

The Gram positive bacteria AI proteins of the invention may be isolated from the native Gram positive bacteria, or they may be recombinantly produced, for instance in a heterologous host. For example, the GAS, GBS, and *S. pneumoniae* antigens of the invention may be isolated from

Streptococcus agalactiae, *S. pyogenes*, *S. pneumoniae*, or they may be recombinantly produced, for instance, in a heterologous host. Preferably, the GBS antigens are prepared using a heterologous host.

The heterologous host may be prokaryotic (*e.g.* a bacterium) or eukaryotic. It is preferably *E.coli*, but other suitable hosts include *Bacillus subtilis, Vibrio cholerae, Salmonella typhi, Salmonella typhimurium, Neisseria lactamica, Neisseria cinerea, Mycobacteria* (*e.g. M.tuberculosis*), *S. gordonii*, *L. lactis*, yeasts, *etc.*

Recombinant production of polypeptides is facilitated by adding a tag protein to the Gram positive bacteria AI sequence to be expressed as a fusion protein comprising the tag protein and the Gram positive bacteria antigen. For example, recombinant production of polypeptides is facilitated by adding a tag protein to the GBS antigen to be expressed as a fusion protein comprising the tag protein and the GBS antigen. Such tag proteins can facilitate purification, detection and stability of the expressed protein. Tag proteins suitable for use in the invention include a polyarginine tag (Arg-tag), polyhistidine tag (His-tag), FLAG-tag, Strep-tag, c-myc-tag, S-tag, calmodulin-binding peptide, cellulose-binding domain, SBP-tag,, chitin-binding domain, glutathione S-transferase-tag (GST), maltose-binding protein, transcription termination anti-terminiantion factor (NusA), *E. coli* thioredoxin (TrxA) and protein disulfide isomerase I (DsbA). Preferred tag proteins include His-tag and GST. A full discussion on the use of tag proteins can be found at Terpe et al., "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems", Appl Microbiol Biotechnol (2003) 60:523 – 533.

After purification, the tag proteins may optionally be removed from the expressed fusion protein, *i.e.*, by specifically tailored enzymatic treatments known in the art. Commonly used proteases include enterokinase, tobacco etch virus (TEV), thrombin, and factor X_a.

GBS polysaccharides

The compositions of the invention may be further improved by including GBS polysaccharides. Preferably, the GBS antigen and the saccharide each contribute to the immunological response in a recipient. The combination is particularly advantageous where the saccharide and polypeptide provide protection from different GBS serotypes.

The combined antigens may be present as a simple combination where separate saccharide and polypeptide antigens are administered together, or they may be present as a conjugated combination, where the saccharide and polypeptide antigens are covalently linked to each other.

Thus the invention provides an immunogenic composition comprising (i) one or more GBS AI proteins and (ii) one or more GBS saccharide antigens. The polypeptide and the polysaccharide may advantageously be covalently linked to each other to form a conjugate.

Between them, the combined polypeptide and saccharide antigens preferably cover (or provide protection from) two or more GBS serotypes (*e.g.* 2, 3, 4, 5, 6, 7, 8 or more serotypes). The serotypes of the polypeptide and saccharide antigens may or may not overlap. For example, the polypeptide might protect against serogroup II or V, while the saccharide protects against either serogroups Ia, Ib, or III. Preferred combinations protect against the following groups of serotypes:

(1) serotypes Ia and Ib, (2) serotypes Ia and II, (3) serotypes Ia and III, (4) serotypes Ia and IV, (5) serotypes Ia and V, (6) serotypes Ia and VI, (7) serotypes Ia and VII, (8) serotypes Ia and VIII, (9) serotypes Ib and II, (10) serotypes Ib and III, (11) serotypes Ib and IV, (12) serotypes Ib and V, (13) serotypes Ib and VI, (14) serotypes Ib and VII, (15) serotypes Ib and VIII, 16) serotypes II and III, (17) serotypes II and IV, (18) serotypes II and V, (19) serotypes II and VI, (20) serotypes II and VII, (21) serotypes II and VII, (22) serotypes III and IV, (23) serotypes III and V, (24) serotypes III and VI, (25) serotypes III and VII, (26) serotypes III and VIII, (27) serotypes IV and V, (28) serotypes IV and VI, (29) serotypes IV and VII, (30) serotypes IV and VIII, (31) serotypes V and VI, (32) serotypes V and VII, (33) serotypes V and VIII, (34) serotypes VI and VII, (35) serotypes VI and VIII, and (36) serotypes VII and VIII.

Still more preferably, the combinations protect against the following groups of serotypes: (1) serotypes Ia and II, (2) serotypes Ia and V, (3) serotypes Ib and II, (4) serotypes Ib and V, (5) serotypes III and II, and (6) serotypes III and V. Most preferably, the combinations protect against serotypes III and V.

Protection against serotypes II and V is preferably provided by polypeptide antigens. Protection against serotypes Ia, Ib and/or III may be polypeptide or saccharide antigens.

Immunogenic compositions and medicaments

Compositions of the invention are preferably immunogenic compositions, and are more preferably vaccine compositions. The pH of the composition is preferably between 6 and 8, preferably about 7. The pH may be maintained by the use of a buffer. The composition may be sterile and/or pyrogen-free. The composition may be isotonic with respect to humans.

Vaccines according to the invention may either be prophylactic (*i.e.* to prevent infection) or therapeutic (*i.e.* to treat infection), but will typically be prophylactic. Accordingly, the invention includes a method for the therapeutic or prophylactic treatment of a Gram positive bacteria infection in an animal susceptible to such gram positive bacterial infection comprising administering to said animal a therapeutic or prophylactic amount of the immunogenic composition of the invention. For example, the invention includes a method for the therapeutic or prophylactic treatment of a *Streptococcus agalactiae*, *S. pyogenes*, or *S. pneumoniae* infection in an animal susceptible to streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of the immunogenic compositions of the invention.

The invention also provides a composition of the invention for use of the compositions described herein as a medicament. The medicament is preferably able to raise an immune response in a mammal (*i.e.* it is an immunogenic composition) and is more preferably a vaccine.

The invention also provides the use of the compositions of the invention in the manufacture of a medicament for raising an immune response in a mammal. The medicament is preferably a vaccine.

The invention also provides kits comprising one or more containers of compositions of the invention. Compositions can be in liquid form or can be lyophilized, as can individual antigens. Suitable containers for the compositions include, for example, bottles, vials, syringes, and test tubes.

Containers can be formed from a variety of materials, including glass or plastic. A container may have a sterile access port (for example, the container may be an intravenous solution bag or a vial having a stopper pierceable by a hypodermic injection needle). The composition may comprise a first component comprising one or more Gram positive bacteria AI proteins. Preferably, the AI proteins are surface AI proteins. Preferably, the AI surface proteins are in an oligomeric or hyperoligomeric form. For example, the first component comprises a combination of GBS antigens or GAS antigens, or *S. pneumoniae* antigens. Preferably said combination includes GBS 80. Preferably GBS 80 is present in an oligomeric or hyperoligomeric form.

The kit can further comprise a second container comprising a pharmaceutically-acceptable buffer, such as phosphate-buffered saline, Ringer's solution, or dextrose solution. It can also contain other materials useful to the end-user, including other buffers, diluents, filters, needles, and syringes. The kit can also comprise a second or third container with another active agent, for example an antibiotic.

The kit can also comprise a package insert containing written instructions for methods of inducing immunity against S agalactiae andor *S. pyogenes* and/or *S pneumoniae* or for treating S agalactiae andor *S. pyogenes* and/or S pneumoniae infections. The package insert can be an unapproved draft package insert or can be a package insert approved by the Food and Drug Administration (FDA) or other regulatory body.

The invention also provides a delivery device pre-filled with the immunogenic compositions of the invention.

The invention also provides a method for raising an immune response in a mammal comprising the step of administering an effective amount of a composition of the invention. The immune response is preferably protective and preferably involves antibodies and/or cell-mediated immunity. This immune response will preferably induce long lasting (*e.g.*, neutralising) antibodies and a cell mediated immunity that can quickly respond upon exposure to one or more GBS and/or GAS and/or *S. pneumoniae* antigens. The method may raise a booster response.

The invention provides a method of neutralizing GBS, GAS, or *S. pneumoniae* infection in a mammal comprising the step of administering to the mammal an effective amount of the immunogenic compositions of the invention, a vaccine of the invention, or antibodies which recognize an immunogenic composition of the invention.

The mammal is preferably a human. Where the vaccine is for prophylactic use, the human is preferably a female (either of child bearing age or a teenager). Alternatively, the human may be elderly (e.g., over the age of 50, 55, 60, 65, 70 or 75) and may have an underlying disease such as diabetes or cancer. Where the vaccine is for therapeutic use, the human is preferably a pregnant female or an elderly adult.

These uses and methods are preferably for the prevention and/or treatment of a disease caused by *Streptococcus agalactiae*, or *S. pyogenes*, or *S. pneumoniae*. The compositions may also be

effective against other streptococcal bacteria. The compositions may also be effective against other Gram positive bacteria.

One way of checking efficacy of therapeutic treatment involves monitoring Gram positive bacterial infection after administration of the composition of the invention. One way of checking efficacy of prophylactic treatment involves monitoring immune responses against the Gram positive bacterial antigens in the compositions of the invention after administration of the composition.

One way of checking efficacy of therapeutic treatment involves monitoring GBS infection after administration of the composition of the invention. One way of checking efficacy of prophylactic treatment involves monitoring immune responses against the GBS antigens in the compositions of the invention after administration of the composition.

A way of assessing the immunogenicity of the component proteins of the immunogenic compositions of the present invention is to express the proteins recombinantly and to screen patient sera or mucosal secretions by immunoblot. A positive reaction between the protein and the patient serum indicates that the patient has previously mounted an immune response to the protein in question- that is, the protein is an immunogen. This method may also be used to identify immunodominant proteins and/or epitopes.

Another way of checking efficacy of therapeutic treatment involves monitoring GBS or GAS or *S pneumoniae* infection after administration of the compositions of the invention. One way of checking efficacy of prophylactic treatment involves monitoring immune responses both systemically (such as monitoring the level of IgG1 and IgG2a production) and mucosally (such as monitoring the level of IgA production) against the GBS and/or GAS and/or *S pneumoniae* antigens in the compositions of the invention after administration of the composition. Typically, GBS and/or GAS and/or S pneumoniae serum specific antibody responses are determined post-immunization but pre-challenge whereas mucosal GBS and/or GAS and/or *S pneumoniae* specific antibody body responses are determined post-immunization and post-challenge.

The vaccine compositions of the present invention can be evaluated in *in vitro* and *in vivo* animal models prior to host, *e.g.*, human, administration.

The efficacy of immunogenic compositions of the invention can also be determined *in vivo* by challenging animal models of GBS and/or GAS and/or S pneumoniae infection, *e.g.*, guinea pigs or mice, with the immunogenic compositions. The immunogenic compositions may or may not be derived from the same serotypes as the challenge serotypes. Preferably the immunnogenic compositions are derivable from the same serotypes as the challenge serotypes. More preferably, the immunogenic composition and/or the challenge serotypes are derivable from the group of GBS and/or GAS and/or *S pneumoniae* serotypes.

In vivo efficacy models include but are not limited to: (i) A murine infection model using human GBS and/or GAS and/or S pneumoniae serotypes; (ii) a murine disease model which is a murine model using a mouse-adapted GBS and/or GAS and/or S pneumoniae strain, such as those

strains outlined above which is particularly virulent in mice and (iii) a primate model using human GBS or GAS or S pneumoniae isolates.

The immune response may be one or both of a TH1 immune response and a TH2 response.

The immune response may be an improved or an enhanced or an altered immune response.

The immune response may be one or both of a systemic and a mucosal immune response.

Preferably the immune response is an enhanced system and/or mucosal response.

An enhanced systemic and/or mucosal immunity is reflected in an enhanced TH1 and/or TH2 immune response. Preferably, the enhanced immune response includes an increase in the production of IgG1 and/or IgG2a and/or IgA

Preferably the mucosal immune response is a TH2 immune response. Preferably, the mucosal immune response includes an increase in the production of IgA.

Activated TH2 cells enhance antibody production and are therefore of value in responding to extracellular infections. Activated TH2 cells may secrete one or more of IL-4, IL-5, IL-6, and IL-10. A TH2 immune response may result in the production of IgG1, IgE, IgA and memory B cells for future protection.

A TH2 immune response may include one or more of an increase in one or more of the cytokines associated with a TH2 immune response (such as IL-4, IL-5, IL-6 and IL-10), or an increase in the production of IgG1, IgE, IgA and memory B cells. Preferably, the enhanced TH2 immune resonse will include an increase in IgG1 production.

A TH1 immune response may include one or more of an increase in CTLs, an increase in one or more of the cytokines associated with a TH1 immune response (such as IL-2, IFNγ, and TNFβ), an increase in activated macrophages, an increase in NK activity, or an increase in the production of IgG2a. Preferably, the enhanced TH1 immune response will include an increase in IgG2a production.

Immunogenic compositions of the invention, in particular, immunogenic composition comprising one or more GAS antigens of the present invention may be used either alone or in combination with other GAS antigens optionally with an immunoregulatory agent capable of eliciting a Th1 and/or Th2 response.

Compositions of the invention will generally be administered directly to a patient. Certain routes may be favored for certain compositons, as resulting in the generation of a more effective immune response, preferably a CMI response, or as being less likely to induce side effects, or as being easier for administration. Direct delivery may be accomplished by parenteral injection (*e.g.* subcutaneously, intraperitoneally, intradermally, intravenously, intramuscularly, or to the interstitial space of a tissue), or by rectal, oral (*e.g.* tablet, spray), vaginal, topical, transdermal (*e.g.* see WO 99/27961) or transcutaneous (*e.g.* see WO 02/074244 and WO 02/064162), intranasal (*e.g.* see WO03/028760), ocular, aural, pulmonary or other mucosal administration.

The invention may be used to elicit systemic and/or mucosal immunity.

In one particularly preferred embodiment, the immunogenic composition comprises one or more GBS or GAS or S pneumoniae antigen(s) which elicits a neutralising antibody response and one or more GBS or GAS or S pneumoniae antigen(s) which elicit a cell mediated immune response. In this way, the neutralising antibody response prevents or inhibits an initial GBS or GAS or S pneumoniae infection while the cell-mediated immune response capable of eliciting an enhanced Th1 cellular response prevents further spreading of the GBS or GAS or S pneumoniae infection. Preferably, the immunogenic composition comprises one or more GBS or GAS or S pneumoniae surface antigens and one or more GBS or GAS or S pneumoniae cytoplasmic antigens. Preferably the immunogenic composition comprises one or more GBS or GAS or S pneumoniae surface antigens or the like and one or other antigens, such as a cytoplasmic antigen capable of eliciting a Th1 cellular response.

Dosage treatment can be a single dose schedule or a multiple dose schedule. Multiple doses may be used in a primary immunisation schedule and/or in a booster immunisation schedule. In a multiple dose schedule the various doses may be given by the same or different routes *e.g.* a parenteral prime and mucosal boost, a mucosal prime and parenteral boost, *etc.*

The compositions of the invention may be prepared in various forms. For example, the compositions may be prepared as injectables, either as liquid solutions or suspensions. Solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared (*e.g.* a lyophilised composition). The composition may be prepared for topical administration *e.g.* as an ointment, cream or powder. The composition may be prepared for oral administration *e.g.* as a tablet or capsule, as a spray, or as a syrup (optionally flavoured). The composition may be prepared for pulmonary administration *e.g.* as an inhaler, using a fine powder or a spray. The composition may be prepared as a suppository or pessary. The composition may be prepared for nasal, aural or ocular administration *e.g.* as drops. The composition may be in kit form, designed such that a combined composition is reconstituted just prior to administration to a patient. Such kits may comprise one or more antigens in liquid form and one or more lyophilised antigens.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of antigen(s), as well as any other components, such as antibiotics, as needed. By 'immunologically effective amount', it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention, or increases a measurable immune response or prevents or reduces a clinical symptom. This amount varies depending upon the health and physical condition of the individual to be treated, age, the taxonomic group of individual to be treated (*e.g.* non-human primate, primate, *etc.*), the capacity of the individual's immune system to synthesise antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

Further Components of the Composition

The composition of the invention will typically, in addition to the components mentioned above, comprise one or more 'pharmaceutically acceptable carriers', which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and lipid aggregates (such as oil droplets or liposomes). Such carriers are well known to those of ordinary skill in the art. The vaccines may also contain diluents, such as water, saline, glycerol, *etc.* Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present. A thorough discussion of pharmaceutically acceptable excipients is available in Gennaro (2000) *Remington: The Science and Practice of Pharmacy.* 20th ed., ISBN: 0683306472.

Adjuvants

Vaccines of the invention may be administered in conjunction with other immunoregulatory agents. In particular, compositions will usually include an adjuvant. Adjuvants for use with the invention include, but are not limited to, one or more of the following set forth below:

A. Mineral Containing Compositions

Mineral containing compositions suitable for use as adjuvants in the invention include mineral salts, such as aluminum salts and calcium salts. The invention includes mineral salts such as hydroxides (*e.g.* oxyhydroxides), phosphates (*e.g.* hydroxyphosphates, orthophosphates), sulfates, *etc.* (*e.g.* see chapters 8 & 9 of *Vaccine Design...* (1995) eds. Powell & Newman. ISBN: 030644867X. Plenum.), or mixtures of different mineral compounds (*e.g.* a mixture of a phosphate and a hydroxide adjuvant, optionally with an excess of the phosphate), with the compounds taking any suitable form (*e.g.* gel, crystalline, amorphous, *etc.*), and with adsorption to the salt(s) being preferred. The mineral containing compositions may also be formulated as a particle of metal salt (WO 00/23105).

Aluminum salts may be included in vaccines of the invention such that the dose of Al^{3+} is between 0.2 and 1.0 mg per dose.

B. Oil-Emulsions

Oil-emulsion compositions suitable for use as adjuvants in the invention include squalene-water emulsions, such as MF59 (5% Squalene, 0.5% Tween 80, and 0.5% Span 85, formulated into submicron particles using a microfluidizer). See WO90/14837. See also, Podda, "The adjuvanted influenza vaccines with novel adjuvants: experience with the MF59-adjuvanted vaccine", Vaccine (2001) 19: 2673-2680; Frey et al., "Comparison of the safety, tolerability, and immunogenicity of a MF59-adjuvanted influenza vaccine and a non-adjuvanted influenza vaccine in non-elderly adults", Vaccine (2003) 21:4234-4237. MF59 is used as the adjuvant in the FLUAD™ influenza virus trivalent subunit vaccine.

Particularly preferred adjuvants for use in the compositions are submicron oil-in-water emulsions. Preferred submicron oil-in-water emulsions for use herein are squalene/water emulsions optionally containing varying amounts of MTP-PE, such as a submicron oil-in-water emulsion containing 4-5% w/v squalene, 0.25-1.0% w/v Tween 80 ™ (polyoxyelthylenesorbitan monooleate), and/or 0.25-1.0% Span 85™ (sorbitan trioleate), and, optionally, N-acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-huydroxyphosphophoryloxy)-ethylamine (MTP-PE), for example, the submicron oil-in-water emulsion known as "MF59" (International Publication No. WO 90/14837; US Patent Nos. 6,299,884 and 6,451,325, incorporated herein by reference in their entireties; and Ott et al., "MF59 -- Design and Evaluation of a Safe and Potent Adjuvant for Human Vaccines" in *Vaccine Design: The Subunit and Adjuvant Approach* (Powell, M.F. and Newman, M.J. eds.) Plenum Press, New York, 1995, pp. 277-296). MF59 contains 4-5% w/v Squalene (e.g. 4.3%), 0.25-0.5% w/v Tween 80™, and 0.5% w/v Span 85™ and optionally contains various amounts of MTP-PE, formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA). For example, MTP-PE may be present in an amount of about 0-500 μg/dose, more preferably 0-250 μg/dose and most preferably, 0-100 μg/dose. As used herein, the term "MF59-0" refers to the above submicron oil-in-water emulsion lacking MTP-PE, while the term MF59-MTP denotes a formulation that contains MTP-PE. For instance, "MF59-100" contains 100 μg MTP-PE per dose, and so on. MF69, another submicron oil-in-water emulsion for use herein, contains 4.3% w/v squalene, 0.25% w/v Tween 80™, and 0.75% w/v Span 85™ and optionally MTP-PE. Yet another submicron oil-in-water emulsion is MF75, also known as SAF, containing 10% squalene, 0.4% Tween 80™, 5% pluronic-blocked polymer L121, and thr-MDP, also microfluidized into a submicron emulsion. MF75-MTP denotes an MF75 formulation that includes MTP, such as from 100-400 μg MTP-PE per dose.

Submicron oil-in-water emulsions, methods of making the same and immunostimulating agents, such as muramyl peptides, for use in the compositions, are described in detail in International Publication No. WO 90/14837 and US Patent Nos. 6,299,884 and 6,45 1,325, incorporated herein by reference in their entireties.

Complete Freund's adjuvant (CFA) and incomplete Freund's adjuvant (IFA) may also be used as adjuvants in the invention.

C. Saponin Formulations

Saponin formulations, may also be used as adjuvants in the invention. Saponins are a heterologous group of sterol glycosides and triterpenoid glycosides that are found in the bark, leaves, stems, roots and even flowers of a wide range of plant species. Saponin from the bark of the *Quillaia saponaria* Molina tree have been widely studied as adjuvants. Saponin can also be commercially obtained from *Smilax ornata* (sarsaprilla), *Gypsophilla paniculata* (brides veil), and *Saponaria officianalis* (soap root). Saponin adjuvant formulations include purified formulations, such as QS21, as well as lipid formulations, such as ISCOMs.

Saponin compositions have been purified using High Performance Thin Layer Chromatography (HP-LC) and Reversed Phase High Performance Liquid Chromatography (RP-HPLC). Specific purified fractions using these techniques have been identified, including QS7, QS17, QS18, QS21, QH-A, QH-B and QH-C. Preferably, the saponin is QS21. A method of production of QS21 is disclosed in US Patent No. 5,057,540. Saponin formulations may also comprise a sterol, such as cholesterol (see WO96/33739).

Combinations of saponins and cholesterols can be used to form unique particles called Immunostimulating Complexs (ISCOMs). ISCOMs typically also include a phospholipid such as phosphatidylethanolamine or phosphatidylcholine. Any known saponin can be used in ISCOMs. Preferably, the ISCOM includes one or more of Quil A, QHA and QHC. ISCOMs are further described in EP0109942, WO 96/11711 and WO 96/33739. Optionally, the ISCOMS may be devoid of additional detergent. See WO 00/07621.

A review of the development of saponin based adjuvants can be found at Barr, et al., "ISCOMs and other saponin based adjuvants", Advanced Drug Delivery Reviews (1998) 32:247-271. See also Sjolander, et al., "Uptake and adjuvant activity of orally delivered saponin and ISCOM vaccines", Advanced Drug Delivery Reviews (1998) 32:321-338.

D. Virosomes and Virus Like Particles (VLPs)

Virosomes and Virus Like Particles (VLPs) can also be used as adjuvants in the invention. These structures generally contain one or more proteins from a virus optionally combined or formulated with a phospholipid. They are generally non-pathogenic, non-replicating and generally do not contain any of the native viral genome. The viral proteins may be recombinantly produced or isolated from whole viruses. These viral proteins suitable for use in virosomes or VLPs include proteins derived from influenza virus (such as HA or NA), Hepatitis B virus (such as core or capsid proteins), Hepatitis E virus, measles virus, Sindbis virus, Rotavirus, Foot-and-Mouth Disease virus, Retrovirus, Norwalk virus, human Papilloma virus, HIV, RNA-phages, Qß-phage (such as coat proteins), GA-phage, fr-phage, AP205 phage, and Ty (such as retrotransposon Ty protein p1). VLPs are discussed further in WO 03/024480, WO 03/024481, and Niikura et al., "Chimeric Recombinant Hepatitis E Virus-Like Particles as an Oral Vaccine Vehicle Presenting Foreign Epitopes", Virology (2002) 293:273-280; Lenz et al., "Papillomarivurs-Like Particles Induce Acute Activation of Dendritic Cells", Journal of Immunology (2001) 5246-5355; Pinto, et al., "Cellular Immune Responses to Human Papillomavirus (HPV)-16 L1 Healthy Volunteers Immunized with Recombinant HPV-16 L1 Virus-Like Particles", Journal of Infectious Diseases (2003) 188:327-338; and Gerber et al., "Human Papillomavrisu Virus-Like Particles Are Efficient Oral Immunogens when Coadministered with Escherichia coli Heat-Labile Entertoxin Mutant R192G or CpG", Journal of Virology (2001) 75(10):4752-4760. Virosomes are discussed further in, for example, Gluck et al., "New Technology Platforms in the Development of Vaccines for the Future", Vaccine (2002) 20:B10 –B16. Immunopotentiating reconstituted influenza virosomes (IRIV) are used as the subunit antigen

delivery system in the intranasal trivalent INFLEXAL™ product {Mischler & Metcalfe (2002) *Vaccine* 20 Suppl 5:B17-23} and the INFLUVAC PLUS™ product.

E. Bacterial or Microbial Derivatives

Adjuvants suitable for use in the invention include bacterial or microbial derivatives such as:

(1) *Non-toxic derivatives of enterobacterial lipopolysaccharide (LPS)*

Such derivatives include Monophosphoryl lipid A (MPL) and 3-O-deacylated MPL (3dMPL). 3dMPL is a mixture of 3 De-O-acylated monophosphoryl lipid A with 4, 5 or 6 acylated chains. A preferred "small particle" form of 3 De-O-acylated monophosphoryl lipid A is disclosed in EP 0 689 454. Such "small particles" of 3dMPL are small enough to be sterile filtered through a 0.22 micron membrane (see EP 0 689 454). Other non-toxic LPS derivatives include monophosphoryl lipid A mimics, such as aminoalkyl glucosaminide phosphate derivatives *e.g.* RC-529. See Johnson *et al.* (1999) *Bioorg Med Chem Lett* 9:2273-2278.

(2) *Lipid A Derivatives*

Lipid A derivatives include derivatives of lipid A from *Escherichia coli* such as OM-174. OM-174 is described for example in Meraldi et al., "OM-174, a New Adjuvant with a Potential for Human Use, Induces a Protective Response with Administered with the Synthetic C-Terminal Fragment 242-310 from the circumsporozoite protein of Plasmodium berghei", Vaccine (2003) 21:2485-2491; and Pajak, et al., "The Adjuvant OM-174 induces both the migration and maturation of murine dendritic cells in vivo", Vaccine (2003) 21:836-842.

(3) *Immunostimulatory oligonucleotides*

Immunostimulatory oligonucleotides suitable for use as adjuvants in the invention include nucleotide sequences containing a CpG motif (a sequence containing an unmethylated cytosine followed by guanosine and linked by a phosphate bond). Bacterial double stranded RNA or oligonucleotides containing palindromic or poly(dG) sequences have also been shown to be immunostimulatory.

The CpG's can include nucleotide modifications/analogs such as phosphorothioate modifications and can be double-stranded or single-stranded. Optionally, the guanosine may be replaced with an analog such as 2'-deoxy-7-deazaguanosine. See Kandimalla, et al., "Divergent synthetic nucleotide motif recognition pattern: design and development of potent immunomodulatory oligodeoxyribonucleotide agents with distinct cytokine induction profiles", Nucleic Acids Research (2003) 31(9): 2393-2400; WO02/26757 and WO99/62923 for examples of possible analog substitutions. The adjuvant effect of CpG oligonucleotides is further discussed in Krieg, "CpG motifs: the active ingredient in bacterial extracts?", Nature Medicine (2003) 9(7): 831-835; McCluskie, et al., "Parenteral and mucosal prime-boost immunization strategies in mice with hepatitis B surface antigen and CpG DNA", FEMS Immunology and Medical Microbiology (2002) 32:179-185; WO98/40100; US Patent No. 6,207,646; US Patent No. 6,239,116 and US Patent No. 6,429,199.

The CpG sequence may be directed to TLR9, such as the motif GTCGTT or TTCGTT. See Kandimalla, et al., "Toll-like receptor 9: modulation of recognition and cytokine induction by novel

synthetic CpG DNAs", Biochemical Society Transactions (2003) 31 (part 3): 654-658. The CpG sequence may be specific for inducing a Th1 immune response, such as a CpG-A ODN, or it may be more specific for inducing a B cell response, such a CpG-B ODN. CpG-A and CpG-B ODNs are discussed in Blackwell, et al., "CpG-A-Induced Monocyte IFN-gamma-Inducible Protein-10 Production is Regulated by Plasmacytoid Dendritic Cell Derived IFN-alpha", J. Immunol. (2003) 170(8):4061-4068; Krieg, "From A to Z on CpG", TRENDS in Immunology (2002) 23(2): 64-65 and WO01/95935. Preferably, the CpG is a CpG-A ODN.

Preferably, the CpG oligonucleotide is constructed so that the 5' end is accessible for receptor recognition. Optionally, two CpG oligonucleotide sequences may be attached at their 3' ends to form "immunomers". See, for example, Kandimalla, et al., "Secondary structures in CpG oligonucleotides affect immunostimulatory activity", BBRC (2003) 306:948-953; Kandimalla, et al., "Toll-like receptor 9: modulation of recognition and cytokine induction by novel synthetic GpG DNAs", Biochemical Society Transactions (2003) 31(part 3):664-658; Bhagat et al., "CpG penta- and hexadeoxyribonucleotides as potent immunomodulatory agents" BBRC (2003) 300:853-861 and WO 03/035836.

(4) *ADP-ribosylating toxins and detoxified derivatives thereof.*

Bacterial ADP-ribosylating toxins and detoxified derivatives thereof may be used as adjuvants in the invention. Preferably, the protein is derived from *E. coli* (i.e., E. coli heat labile enterotoxin "LT), cholera ("CT"), or pertussis ("PT"). The use of detoxified ADP-ribosylating toxins as mucosal adjuvants is described in WO95/17211 and as parenteral adjuvants in WO98/42375. Preferably, the adjuvant is a detoxified LT mutant such as LT-K63, LT-R72, and LTR192G. The use of ADP-ribosylating toxins and detoxified derivaties thereof, particularly LT-K63 and LT-R72, as adjuvants can be found in the following references, each of which is specifically incorporated by reference herein in their entirety: Beignon, et al., "The LTR72 Mutant of Heat-Labile Enterotoxin of Escherichia coli Enahnces the Ability of Peptide Antigens to Elicit CD4+ T Cells and Secrete Gamma Interferon after Coapplication onto Bare Skin", Infection and Immunity (2002) 70(6):3012-3019; Pizza, et al., "Mucosal vaccines: non toxic derivatives of LT and CT as mucosal adjuvants", Vaccine (2001) 19:2534-2541; Pizza, et al., "LTK63 and LTR72, two mucosal adjuvants ready for clinical trials" Int. J. Med. Microbiol (2000) 290(4-5):455-461; Scharton-Kersten et al., "Transcutaneous Immunization with Bacterial ADP-Ribosylating Exotoxins, Subunits and Unrelated Adjuvants", Infection and Immunity (2000) 68(9):5306-5313; Ryan et al., "Mutants of Escherichia coli Heat-Labile Toxin Act as Effective Mucosal Adjuvants for Nasal Delivery of an Acellular Pertussis Vaccine: Differential Effects of the Nontoxic AB Complex and Enzyme Activity on Th1 and Th2 Cells" Infection and Immunity (1999) 67(12):6270-6280; Partidos et al., "Heat-labile enterotoxin of Escherichia coli and its site-directed mutant LTK63 enhance the proliferative and cytotoxic T-cell responses to intranasally co-immunized synthetic peptides", Immunol. Lett. (1999) 67(3):209-216; Peppoloni et al., "Mutants of the Escherichia coli heat-labile enterotoxin as safe and strong adjuvants for intranasal delivery of vaccines", Vaccines (2003) 2(2):285-293; and Pine et al., (2002) "Intranasal

immunization with influenza vaccine and a detoxified mutant of heat labile enterotoxin from Escherichia coli (LTK63)" J. Control Release (2002) 85(1-3):263-270. Numerical reference for amino acid substitutions is preferably based on the alignments of the A and B subunits of ADP-ribosylating toxins set forth in Domenighini et al., Mol. Microbiol (1995) 15(6):1165-1167, specifically incorporated herein by reference in its entirety.

F. Bioadhesives and Mucoadhesives

Bioadhesives and mucoadhesives may also be used as adjuvants in the invention. Suitable bioadhesives include esterified hyaluronic acid microspheres (Singh *et al.* (2001) *J. Cont. Rele.* 70:267-276) or mucoadhesives such as cross-linked derivatives of poly(acrylic acid), polyvinyl alcohol, polyvinyl pyrollidone, polysaccharides and carboxymethylcellulose. Chitosan and derivatives thereof may also be used as adjuvants in the invention. E.g. WO99/27960.

G. Microparticles

Microparticles may also be used as adjuvants in the invention. Microparticles (*i.e.* a particle of ~100nm to ~150μm in diameter, more preferably ~200nm to ~30μm in diameter, and most preferably ~500nm to ~10μm in diameter) formed from materials that are biodegradable and non-toxic (*e.g.* a poly(α-hydroxy acid), a polyhydroxybutyric acid, a polyorthoester, a polyanhydride, a polycaprolactone, *etc.*), with poly(lactide-co-glycolide) are preferred, optionally treated to have a negatively-charged surface (*e.g.* with SDS) or a positively-charged surface (*e.g.* with a cationic detergent, such as CTAB).

H. Liposomes

Examples of liposome formulations suitable for use as adjuvants are described in US Patent No. 6,090,406, US Patent No. 5,916,588, and EP 0 626 169.

I. Polyoxyethylene ether and Polyoxyethylene Ester Formulations

Adjuvants suitable for use in the invention include polyoxyethylene ethers and polyoxyethylene esters. WO99/52549. Such formulations further include polyoxyethylene sorbitan ester surfactants in combination with an octoxynol (WO01/21207) as well as polyoxyethylene alkyl ethers or ester surfactants in combination with at least one additional non-ionic surfactant such as an octoxynol (WO 01/21152).

Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether (laureth 9), polyoxyethylene-9-steoryl ether, polyoxytheylene-8-steoryl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether.

J. Polyphosphazene (PCPP)

PCPP formulations are described, for example, in Andrianov et al., "Preparation of hydrogel microspheres by coacervation of aqueous polyphophazene solutions", Biomaterials (1998) 19(1-3):109-115 and Payne et al., "Protein Release from Polyphosphazene Matrices", Adv. Drug. Delivery Review (1998) 31(3):185-196.

K. Muramyl peptides

Examples of muramyl peptides suitable for use as adjuvants in the invention include N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-l-alanyl-d-isoglutamine (nor-MDP), and N-acetylmuramyl-l-alanyl-d-isoglutaminyl-l-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine MTP-PE).

L. Imidazoquinolone Compounds.

Examples of imidazoquinolone compounds suitable for use adjuvants in the invention include Imiquamod and its homologues, described further in Stanley, "Imiquimod and the imidazoquinolones: mechanism of action and therapeutic potential" Clin Exp Dermatol (2002) 27(7):571-577 and Jones, "Resiquimod 3M", Curr Opin Investig Drugs (2003) 4(2):214-218.

The invention may also comprise combinations of aspects of one or more of the adjuvants identified above. For example, the following adjuvant compositions may be used in the invention:

(1) a saponin and an oil-in-water emulsion (WO 99/11241);

(2) a saponin (e.g.., QS21) + a non-toxic LPS derivative (e.g. 3dMPL) (see WO 94/00153);

(3) a saponin (e.g.., QS21) + a non-toxic LPS derivative (e.g. 3dMPL) + a cholesterol;

(4) a saponin (*e.g.* QS21) + 3dMPL + IL-12 (optionally + a sterol) (WO 98/57659);

(5) combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions (See European patent applications 0835318, 0735898 and 0761231);

(6) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-block polymer L121, and thr-MDP, either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion.

(7) Ribi™ adjuvant system (RAS), (Ribi Immunochem) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™);

(8) one or more mineral salts (such as an aluminum salt) + a non-toxic derivative of LPS (such as 3dPML).

(9) one or more mineral salts (such as an aluminum salt) + an immunostimulatory oligonucleotide (such as a nucleotide sequence including a CpG motif). Combination No. (9) is a preferred adjuvant combination.

M. Human Immunomodulators

Human immunomodulators suitable for use as adjuvants in the invention include cytokines, such as interleukins (*e.g.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, *etc.*), interferons (*e.g.* interferon-γ), macrophage colony stimulating factor, and tumor necrosis factor.

Aluminum salts and MF59 are preferred adjuvants for use with injectable influenza vaccines. Bacterial toxins and bioadhesives are preferred adjuvants for use with mucosally-delivered vaccines, such as nasal vaccines.

The immunogenic compositions of the present invention may be administed in combination with an antibiotic treatment regime. In one embodiment, the antibiotic is administered prior to administration of the antigen of the invention or the composition comprising the one or more of the antigens of the invention.

In another embodiment, the antibiotic is administered subsequent to the adminstration of the one or more antigens of the invention or the composition comprising the one or more antigens of the invention. Examples of antibiotics suitable for use in the treatment of the Steptococcal infections of the invention include but are not limited to penicillin or a derivative thereof or clindamycin or the like.

Further antigens

The compositions of the invention may further comprise one or more additional Gram positive bacterial antigens which are not associated with an AI. Preferably, the Gram positive bacterial antigens that are not associated with an AI can provide protection across more than one serotype or strain isolate. For example, a first non-AI antigen, in which the first non-AI antigen is at least 90% (*i.e.*, at least 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) homologous to the amino acid sequence of a second non-AI antigen, wherein the first and the second non-AI antigen are derived from the genomes of different serotypes of a Gram positive bacteria, may be further included in the compositions. The first non-AI antigen may also be homologous to the amino acid sequence of a third non-AI antigen, such that the first non-AI antigen, the second non-AI antigen, and the third non-AI antigen are derived from the genomes of different serotypes of a Gram positive bacteria. The first non-AI antigen may also be homologous to the amino acid sequence of a fourth non-AI antigen, such that the first non-AI antigen, the second non-AI antigen, the third non-AI antigen, and the fourth non-AI antigen are derived from the genomes of different serotypes of a Gram positive bacteria.

The first non-AI antigen may be GBS 322. The amino acid sequence of GBS 322 across GBS strains from serotypes Ia, Ib, II, III, V, and VIII is greater than 90%. Alternatively, the first non-AI antigen may be GBS 276. The amino acid sequence of GBS 276 across GBS strain from serotypes Ia, Ib, II, III, V, and VIII is greater than 90%. Table 13 provides the percent amino acid sequence identity of GBS 322 and GBS 276 across different GBS strains and serotypes.

Table 13: Conservation of GBS 322 and GBS 276 amino acid sequences

Serotype	Strains	GBS 322		GBS 276	
		cGH	%AA identity	cGH	%AA identity
Ia	**090**	+	98.60	+	97.90
	A909	+	98.30	+	97.90
	515	+	98.80	+	97.50
	DK1	+		+	
	DK8	+		+	
	Davis	+		+	
Ib	**7357b**	+		+	
	H36B	+	98.30	+	97.80
II	**18RS21**	+	100.00	+	99.90
	DK21	+		+	

Serotype	Strains	GBS 322		GBS 276	
		cGH	%AA identity	cGH	%AA identity
III	NEM316	+	100.00	+	97.00
	COH31	+		+	
	D136	+		+	
	M732	+	98.00	+	100.00
	COH1	+	98.30	+	100.00
	M781	+	98.30	+	99.60
No type	CJB110	+	98.60	+	97.90
	1169NT	+	97.40	+	97.90
V	CJB111	+	100.00	+	
	2603	+	100.00	+	100.00
VIII	JM130013	+	100.00	+	97.90
	SMU014	+		+	
total		22/22	98.28+/-0.4	22/22	98.44 +/-1.094

As an example, inclusion of a non-AI protein, GBS 322, in combination with AI antigens GBS 67, GBS 80, and GBS 104 provided protection to newborn mice in an active maternal immunization assay.

Table 14: Active maternal immunization assay for a combination of fragments from GBS 322, GBS 80, GBS 104, and GBS 67

		FACS (Δ Mean)			MIX=322+80+104+67		PBS	
GBS strains	Type	GBS 80	GBS 67	GBS 322	alive/treated	% protection	alive/treated	% protection
515	Ia	0	409	227	39/40	97	6/40	15
7357b-	Ib	91	316	102	19/30	63	1/30	3
DK21	II	0	331	416	25/34	73	17/48	35
5401	II	170	618	135	35/40	87	3/37	8
3050	II	43	460	188	48/48	100	1/30	3
COH1	III	305	0	130	36/36	100	7/40	17
M781	III	65	0	224	30/40	75	4/39	10
2603	V	125	105	313	27/33	82	10/35	28
CJB111	V	370	481	63	25/28	89	4/46	9
JM9130013	VIII	597	83	143	37/39	95	5/40	12
JMU071	VIII	556	79	170	44/50	88	18/50	36
NT1169	NT	0	443	213	12/32	37	11/35	31

In fact, the non-AI GBS 322 antigen may itself provide protection to newborn mice in an active maternal immunization assay.

PCT/US05/27239

Table 16: Active maternal immunization assay for each of GBS 80 and GBS 322 antigens

		GBS 80			GBS 322		
		FACS	Protection (% survival)		FACS	Protection (% survival)	
GBS strains	Type	Δ Mean	antigen	ctrl-	Δ Mean	antigen	ctrl-
CJB111	V	370	72 %	40%	63	57%	40%
COH1	III	305	76 %	10%	130	3%	10%
2603	V	82	22 %	34%	313	83 %	34%
7357b-	Ib	91	36%	34%	102	43%	34%
18RS21	II	0	15%	24%	268	84 %	24%
DK21	II	0	10%	21%	416	67 %	25%
A909	Ia	0	0%	14%			
090	Ia	0	0%	0%			
H36B	Ib				105	34%	32%

Thus, inclusion of a non-AI protein in an immunogenic composition of the invention may provide increased protection a mammal.

The immunogenic compositions comprising *S. pneumonaie* AI polypeptides may further secondary SP protein antigens which include (a) any of the SP protein antigens disclosed in WO 02/077021 or U.S. provisional application _____, filed April 20, 2005 (Attorney Docket Number 002441.00154), (2) immunogenic portions of the antigens comprising at least 7 contiguous amino acids, (3) proteins comprising amino acid sequences which retain immunogenicity and which are at least 95% identical to these SP protein antigens (*e.g.*, 95%, 96%, 97%, 98%, 99%, or 99.5% identical), and (4) fusion proteins, including hybrid SP protein antigens, comprising (1)-(3).

Alternatively, the invention may include an immunogenic composition comprising a first and a second Gram positive bacteria non-AI protein, wherein the polynucleotide sequence encoding the sequence of the first non-AI protein is less than 90% (i.e., less than 90, 88, 86, 84, 82, 81, 78, 76, 74, 72, 70, 65, 60, 55, 50, 45, 40, 35, or 30 percent) homologous than the corresponding sequence in the genome of the second non-AI protein.

The compositions of the invention may further comprise one or more additional non-Gram positive bacterial antigens, including additional bacterial, viral or parasitic antigens. The compositions of the invention may further comprise one or more additional non-GBS antigens, including additional bacterial, viral or parasitic antigens.

In another embodiment, the GBS antigen combinations of the invention are combined with one or more additional, non-GBS antigens suitable for use in a vaccine designed to protect elderly or immunocomprised individuals. For example, the GBS antigen combinations may be combined with an antigen derived from the group consisting of *Enterococcus faecalis, Staphylococcus aureus, Staphylococcus epidermis, Pseudomonas aeruginosa, Legionella pneumophila, Listeria monocytogenes, Neisseria meningitides,* influenza, and Parainfluenza virus ('PIV').

Where a saccharide or carbohydrate antigen is used, it is preferably conjugated to a carrier protein in order to enhance immunogenicity {*e.g.* Ramsay *et al.* (2001) *Lancet* 357(9251):195-196; Lindberg (1999) *Vaccine* 17 Suppl 2:S28-36; Buttery & Moxon (2000) *J R Coll Physicians Lond* 34:163-168; Ahmad & Chapnick (1999) *Infect Dis Clin North Am* 13:113-133, vii.; Goldblatt (1998) *J. Med. Microbiol.* 47:563-567; European patent 0 477 508; US Patent No. 5,306,492; International patent application WO98/42721; *Conjugate Vaccines* (eds. Cruse *et al.*) ISBN 3805549326, particularly vol. 10:48-114; and Hermanson (1996) *Bioconjugate Techniques* ISBN: 0123423368 or 012342335X}. Preferred carrier proteins are bacterial toxins or toxoids, such as diphtheria or tetanus toxoids. The CRM_{197} diphtheria toxoid is particularly preferred {*Research Disclosure*, 453077 (Jan 2002)}. Other carrier polypeptides include the *N.meningitidis* outer membrane protein (EP-A-0372501), synthetic peptides (EP-A-0378881; EP-A-0427347), heat shock proteins (WO 93/17712; WO 94/03208), pertussis proteins (WO 98/58668; EP A 0471177), protein D from *H.influenzae* (WO 00/56360), cytokines (WO 91/01146), lymphokines, hormones, growth factors, toxin A or B from *C.difficile* (WO00/61761), iron-uptake proteins (WO01/72337), *etc.* Where a mixture comprises capsular saccharides from both serogroups A and C, it may be preferred that the ratio (w/w) of MenA saccharide:MenC saccharide is greater than 1 (*e.g.* 2:1, 3:1, 4:1, 5:1, 10:1 or higher). Different saccharides can be conjugated to the same or different type of carrier protein. Any suitable conjugation reaction can be used, with any suitable linker where necessary.

Toxic protein antigens may be detoxified where necessary *e.g.* detoxification of pertussis toxin by chemical and/or genetic means.

Where a diphtheria antigen is included in the composition it is preferred also to include tetanus antigen and pertussis antigens. Similarly, where a tetanus antigen is included it is preferred also to include diphtheria and pertussis antigens. Similarly, where a pertussis antigen is included it is preferred also to include diphtheria and tetanus antigens.

Antigens in the composition will typically be present at a concentration of at least 1μg/ml each. In general, the concentration of any given antigen will be sufficient to elicit an immune response against that antigen.

As an alternative to using protein antigens in the composition of the invention, nucleic acid encoding the antigen may be used {*e.g.* refs. Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; Scott-Taylor & Dalgleish (2000) *Expert Opin Investig Drugs* 9:471-480; Apostolopoulos & Plebanski (2000) *Curr Opin Mol Ther* 2:441-447; Ilan (1999) *Curr Opin Mol Ther* 1:116-120; Dubensky *et al.* (2000) *Mol Med* 6:723-732; Robinson & Pertmer (2000) *Adv Virus Res* 55:1-74; Donnelly *et al.* (2000) *Am J Respir Crit Care Med* 162(4 Pt 2):S190-193; and Davis (1999) *Mt. Sinai J. Med.* 66:84-90}. Protein components of the compositions of the invention may thus be replaced by nucleic acid (preferably DNA *e.g.* in the form of a plasmid) that encodes the protein.

Definitions PCT/US05/27239

The term "comprising" means "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The term "about" in relation to a numerical value *x* means, for example, $x \pm 10\%$.

References to a percentage sequence identity between two amino acid sequences means that, when aligned, that percentage of amino acids are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in section 7.7.18 of *Current Protocols in Molecular Biology* (F.M. Ausubel *et al.*, eds., 1987) Supplement 30. A preferred alignment is determined by the Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 2, BLOSUM matrix of 62. The Smith-Waterman homology search algorithm is disclosed in Smith & Waterman (1981) *Adv. Appl. Math.* 2: 482-489.

The invention is further illustrated, without limitation, by the following examples.

EXAMPLE 1: Binding of an Adhesin Island surface protein, GBS 80, to Fibrinogen and Fibronectin.

This example demonstrates that an Adhesin Island surface protein, GBS 80 can bind to fibrinogen and fibronectin.

An enzyme-linked immunosorbent assay (ELISA) was used to analyse the *in vitro* binding ability of recombinant GBS 80 to immobilized extra-cellular matrix (ECM) proteins but not to bovine serum albumin (BSA). Microtiter plates were coated with ECM proteins (fibrinogen, fibronectin, laminin, collagen type IV) and binding assessed by adding varying concentrations of a recombinant form of GBS 80, over-expressed and purified from *E. coli* (FIGURE 5A). Plates were then incubated sequentially with a) mouse anti-GBS 80 primary antibody; b) rabbit anti-mouse AP-conjugated secondary antibody; c) pNPP colorimetric substrate. Relative binding was measured by monitoring absorbance at 405 nm, using 595 nm as a reference wavelength. Figure 5b shows binding of recombinant GBS 80 to immobilized ECM proteins (1 μg) as a function of concentration of GBS 80. BSA was used as a negative control. Data points represent the means of OD_{405} values ± standard deviation for 3 wells.

Binding of GBS 80 to the tested ECM proteins was found to be concentration dependent and exhibited saturation kinetics. As is also evident from FIGURE 5, binding of GBS 80 to fibronectin and fibrinogen was greater than binding to laminin and collagen type IV at all the concentrations tested.

EXAMPLE 2: GBS 80 is required for surface localization of GBS 104.

This example demonstrates that co-expression of GBS 80 is required for surface localization of GBS 104.

The polycistronic nature of the Adhesin Island I mRNA was investigated through reverse transcriptase-PCR (RT-PCR) analysis employing primers designed to detect transcripts arising from contiguous genes. Total RNA was isolated from GBS cultures grown to an optical density at 600 nm

(OD_{600}) of 0.3 in THB (Todd-Hewitt broth) by the RNeasy Total RNA isolation method (Qiagen) according to the manufacturer's instructions. The absence of contaminating chromosomal DNA was confirmed by failure of the gene amplification reactions to generate a product detectable by agarose gel electrophoresis, in the absence of reverse transcriptase. RT-PCR analysis was performed with the Access RT-PCR system (Promega) according to the manufacturer's instructions, employing PCR cycling temperatures of 60°C for annealing and 70°C for extension. Amplification products were visualized alongside 100-bp DNA markers in 2% agarose gels after ethidium bromide staining.

FIGURE 5 shows that all the genes are co-transcribed as an operon. A schematic of the AI-1 operon is shown above the agarose gel analysis of the RT-PCR products. Large rectangular arrows indicate the predicted transcript direction. Primer pairs were selected such as "1-4" cross the 3'finish-5'start of successive genes and overlap each gene by at least 200 bp. Additionally, "1" crosses a putative rho-independent transcriptional terminator. "5" is an internal GBS 80 control and "6" is an unrelated control from a highly expressed gene. Lanes: "a": RNA plus RTase enzyme; "b" RNA without RTase; "c": genomic DNA control.

In the effort to elucidate the functions of the AI-1 proteins, in frame deletions of all of the genes within the operon have been constructed and the resulting mutants characterized with respect to surface exposure of the encoded antigens (see FIGURE 8).

Each in-frame deletion mutation was constructed by splice overlap extension PCR (SOE-PCR) essentially as decribed by Horton et al. [Horton R. M., Z. L. Cai, S. N. Ho, L. R. Pease (1990) Biotechniques 8:528-35] using suitable primers and cloned into the temperature sensitive shuttle vector pJRS233 to replace the wild type copy by allelic exchange [Perez-Casal, J., J. A. Price, et al. (1993) Mol Microbiol 8(5): 809-19.]. All plasmid constructions utilized standard molecular biology techniques, and the identities of DNA fragments generated by PCR were verified by sequencing. Following SOE-PCR, the resulting mutant DNA fragments were digested with XhoI and EcoRI, and ligated into a similarly digested pJRS233. The resuting vectors were introduced by electroporation into the chromosome of 2603 and COH1 GBS strains in a three-step process, essentially as described in Framson et al. [Framson, P. E., A. Nittayajarn, J. Merry, P. Youngman, and C. E. Rubens. (1997) Appl. Environ. Microbiol. 63(9):3539-47]. Briefly, the vector pJRS233 contains an *erm* gene encoding erythromycin resistance and a temperature-sensitive gram-positive replicon that is active at 30°C but not at 37°C. Initially, the constructs are electroporated into GBS electro-competent cells prepared as described by Frameson et al., and transformants containing free plasmid are selected by their ability to grow at 30°C on Todd-Hewitt Broth (THB) agar plates containing 1 μg/ml erythromycin. The second step includes a selection step for strains in which the plasmid has integrated into the chromosome via a single recombination event over the homologous plasmid insert and chromosome sequence by their ability to grow at 37°C on THB agar medium containing 1 mg/ml erythromycin. In the third step, GBS cells containing the plasmid integrated within the chromosome (integrants) are serially passed in broth culture in the absence of antibiotics at 30°C. Plasmid excision